ESTIMATING THE DISCRIMINATIVE POWER OF TIME VARYING FEATURES FOR EEG BMI

A Thesis
Presented to
The Academic Faculty

by

Rudolph L. Mappus IV
cmappus@gatech.edu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Interactive Computing, College of Computing

Georgia Institute of Technology
December 2009

ESTIMATING THE DISCRIMINATIVE POWER OF TIME VARYING FEATURES FOR EEG BMI

Approved by:

Professor Charles L. Isbell Jr., Advisor
Interactive Computing, College of Computing
Georgia Institute of Technology

Professor Alexander Gray
Computer Science and Engineering, College of Computing
Georgia Institute of Technology

Professor Melody M. Jackson
Interactive Computing, College of Computing
Georgia Institute of Technology

Professor Thad Starner
Interactive Computing, College of Computing
Georgia Institute of Technology

Professor Paul M. Corballis
Department of Psychology
Georgia Institute of Technology

Date Approved: 16 November 2009

For Amelia.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Charles Isbell, for allowing me the freedom to follow this investigation. I would also like to thank my committee: Dr. Thad Starner, Dr. Alex Gray, Dr. Melody Moore Jackson, and Dr. Paul Corballis for their guidance and support every step of the way. Each member has played a significant role in overseeing this work; I am thankful for each of their efforts.

I would also like to thank a number of students whose helpful suggestions have made a positive impact on this work: Peng Zang, Sooraj Bhat, Arya Irani, Tracy Westeyn, Nate Parks, Matt Hilimire, and Nishant Mehta.

I want to acknowledge my wife Elena, whose endless love and support has been a constant source of encouragement and motivation. Your support has in no small way made this work possible. I want to thank you for letting me pursue this dream, for your gentle understanding, and for your unwavering advocacy during this time. This acknowledgement can not express how fortunate I have been to have you along this journey. I love you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Research in brain-machine interfaces requires integration of a number of separate research topics. Indeed, the ability to use observations of brain activity as a channel of interaction is a goal of a number of research communities. Progress in robust brain-machine interfaces (BMIs) requires a number of specializations: brain function, recording techniques for brain activity, computational models of brain activity, feature selection and classification, and feedback mechanisms. Modern BMIs are the integration of these desiderata.

Today, there is a growing need for robust BMIs as an assistive technology. In 2005, it was estimated that 3.17 million Americans were currently living with disabilities resulting from traumatic brain injury (TBI) (102; 71). Traumatic brain injury is the leading cause of disability in children and adults from ages 1 to 44 (1). Many of these people require alternative interaction technologies for both therapy and quality of life. People suffering from degenerative diseases are living longer and require more sophisticated interaction channels: fifty percent of people with Amyotrophic Lateral Sclerosis (ALS) live at least three years after diagnosis, twenty percent at least five years, and ten percent at least ten years (39).

We make the following claim: *Sparse regularization improves components analysis in noisy, overcomplete environments; a psychophysiological analysis of mental rotation shows its applicability to BMIs. The combination of these approaches enables more robust BMI.* This dissertation is an explanation and elaboration of these concepts and serves as evidence for our claims.

Current research in brain-machine interfaces is progressing from answering questions of effectiveness to questions of efficiency: what brain-machine interface (BMI) approaches facilitate robust interaction? Currently, robust interaction with BMIs is limited by problems of initiating and stopping interaction as well as the presence of artifacts and noise in sensor data. Increased efficiency for BMIs means greater accessibility for different populations. For

the disabled population, more ways of indicating intention means greater accessibility for a greater range of impairments. For the general population, new methods of interaction allow for tighter, closed-loop biofeedback mechanisms, with applications such as simple desktop task control, game interaction, and remote robotic control.

Portable sensing arrays such as electroencephalography (EEG) are commonly used for applying neural signals to near real-time control tasks, because they offer minimally invasive sensing arrays for observing neural activity. EEG is particularly able to observe electrical activity of neural cells in response to stimuli with good temporal resolution. These event related potentials (ERPs) are the target signals used by BMIs, and one objective of ERP inference is to be able to identify target activity of a single trial. However, EEG is spatially sparse and sensitive to electrical noise, therefore robust inference requires effective methods for removing artifacts and segmenting target signals. Being able to factor noise artifacts allows us to recover features that better represent the underlying functional processes within the brain. ***The primary objective of this work is to improve signal classification of ERP data by improving noise factoring methods and by discovering novel ERP patterns.***

CHAPTER I

INTRODUCTION

Brain-machine interfaces (BMI) are systems that detect small changes in brain activity and infer user intent from those changes in the context of a control task. These interfaces represent an emerging method of human-machine interaction, and a key challenge lies in processing the detected changes. Issues of effectiveness and efficiency lie in evoking and recognizing patterns in the data. The ability to do this is hampered by a number of factors such as noise from sensors and the environment as well as subject artifacts as subtle as blinking and involuntary like perspiration.

The focus of research in BMIs is changing from answering questions of effectiveness to questions of efficiency: what approaches facilitate robust interaction? Increased efficiency for BMIs means greater accessibility for different populations. For the disabled population, more ways of indicating intention means greater accessibility for a greater range of impairments. For the general population, new methods of interaction allow for tighter, closed-loop biofeedback mechanisms, with applications such as simple desktop task control, game interaction, and remote robotic control. BMIs may also be used in pathology and intervention assessment providing a quantitative measure previously unavailable for clinicians for estimating both the progressive state of degenerative diseases and therapy effectiveness. An important area of application for this work is that of pathology and intervention assessment. The ability to use relatively low-cost sensor arrays in clinical and at-home environments provides a new set of tools for monitoring and assessing pathologies and interventions for disabilities. More effective at-home rehabilitation methods relieve clinical case loads and allow for more effective rehabilitation both from a motivational and sustainable perspective. Reliable activity measures give anesthesiologists tools for measuring depth of sedation, improving recovery time from surgery.

We approach the question of efficiency from several sides of the problem and make

contributions to each in the context of brain-machine interfaces. First, we develop two methods for selecting features to improve independent components analysis. Second, we present results of an empirical study of ERPs for mental rotation tasks. Third, we present results of an empirical study of ERPs for symmetry perception. Finally we demonstrate the application of sparse feature selection to classification of single trial ERPs using the mental rotation and symmetry study data.

1.1 Specific Aims

We propose a set of specific aims to demonstrate ERP signal classification improvement for BMIs:

1. **Demonstrate detection of mental rotation tasks in ERP data.** We hypothesize that networks that participate in mental rotation tasks are *orientation selective* meaning that networks selectively respond to angles of rotation. The ERP differences of orientation selective networks are differentiable from the ERP of language production tasks. Language production typically activates spatially separate regions of the brain and represents a contrasting activation pattern evoked using the same stimuli. We validate the hypothesis with a set of experiments measuring ERP differences between language production and rotation tasks.

2. **Demonstrate the detection of perceived symmetry in ERP data.** We hypothesize that perceived symmetry is represented in the early ERP response (200ms). By manipulating aspects of approximate symmetry, we hypothesize that ERP differences between correct and incorrect responses will correlate with the ERP of correct symmetry and asymmetry conditions. We validate the hypothesis with an experiment measuring ERP differences to polygon stimuli with systematic manipulations of approximate symmetry. The implications of this work are that symmetry manipulations inform graphical user interface design.

3. **Define feature selection methods for improving interpretability of components generated using independent components analysis methods.** We

present two methods that explore noise factoring through dimensionality reduction and regularization in the context of independent components analysis. We hypothesize that dimensionality reduction and regularization of noise elements produce components that are more effective for classification than ICA alone. We validate the method with three experiments designed to test the effectiveness of separation in noisy and overcomplete situations. We also validate the method against the datasets generated by the first two aims.

This dissertation analyzes two sets of ERP data to demonstrate feature recovery using sparse regularization. The data collected in our mental rotation and symmetry studies serve as validation sets for sparse regularization. The long-term goal of using sparse regularization feature selection methods with ERP data is to support single-trial classification of ERPs, an area of active research in the BMI community.

1.2 Feature selection

A primary analysis tool of ERP data is independent components analysis. Finding the underlying causes in ERP data gives researchers insight into finding response differences between experimental and control data. Unfortunately, ICA methods make an assumption of completeness: that the number of sources is equal to the number of sensors. In ERP data, this is rarely the case. Often sensor arrays are sparse and subject to noise from the environment as well as the subject. Further, in dense array conditions, it can be difficult to interpret the number of sources recovered. Being able to recover sources that are less noisy and more easily interpreted enables better analysis of ERP data. We extend independent component analysis methods for selecting time varying features. The goal is to select components and features that best represent target signals while maximizing interpretability of components. We incorporate sparse regularization into the optimization objective functions of several existing ICA methods. This represents an analysis tool for ICA users, particularly with ERP data. Embedded regularization formulations of source separation also represents an important contribution to machine learning research, simultaneously optimizing for statistical independence with sparse regularization.

1.3 ERP experiments

We present results of two ERP experiments. The first study examines the ability to detect subtle differences in activation as a result of orientation selective networks to mental rotation tasks. The results of the study are relevant to BMIs, offering a more detailed analysis of mental rotation tasks and suggesting that mental rotation alone can be used for control. The second study involves approximate symmetry perception. The study data serves as an additional validation set, demonstrating the effectiveness of our feature selection method using ERP data from completely different study.

1.3.1 Mental rotation

In pyschophysiological research, a converging set of evidence is building that mental rotation tasks are handled in the parietal lobe; specifically about the intraparietal sulcus (IPS). The question we ask here is can we detect different activation patterns in IPS for mental rotation in ERP data? Being able to detect these differences would make mental rotation tasks an attractive objective task for BMI. Mental rotation is a rare event task (*i.e.* rotation is not activated by passive stimuli). The experimental results presented in this dissertation test the ability to detect these differences in ERP data using two-dimensional letter data. The first experiment contrasts letter rotation tasks with language production tasks. The second experiment compares rotation tasks with explicit judgment to rotation with covert judgment.

1.3.2 Symmetry

Symmetry is a geometric property of many objects in the visual environment both natural and man-made. Many objects are not exact-symmetric but are approximately symmetric. Despite approximate symmetry, the human visual system detects symmetry readily even through occlusion and oblique angles. How does the visual system represent information of symmetry? Understanding symmetry information within the visual cortex enables more effective interface design by leveraging what humans are better able to detect. The aim of

this study is to determine the effects of certain types of approximate symmetry manipulations on participants' ability to assess symmetry. The findings indicate that symmetry decisions are made early in perception. The study data serves as validation for our sparse regularization feature selection method.

1.4 Summary of Conclusions

The results presented here on sparse regularization and mental rotation support two claims about brain-machine interfaces: improved interface efficiency is affected by methods that account for noise and a psychophysiological approach to mental task selection expands the kinds of mental tasks that can be used in brain-machine interfaces. Advances in both areas improves accessibility of BMIs and changes the way people think about applying BMIs to interaction problems.

1.5 Terminology

This work is broadly relevant to the fields of artificial intelligence, psychophysiology, and human-computer interaction. Let us define some terms used throughout this dissertation that may hold different interpretations from each of these fields.

BMIs process signals of brain activity and produce control signals in near real-time in the service of a control task. For ERP based BMI, response is near real-time rather than real-time due to the fact that there is inherent latency in ERP signal acquisition, processing, and classification. Single-trial classification in ERP refers to methods capable of classifying activity based on no training data.

We refer to data using the traditional machine learning formulation where X is a set of instances: $x_n \in X$, and n represents the number of instances. Each instance is a fixed length vector of features $x_i, i : 1..p$ where p is the dimensionality of the data. A discrete vector of corresponding labels y_n attributes a class to each instance. A classifier is a method that recovers a mapping of features to labels. Independent components analysis is a method that produces a transform that produces an alternative set of features for each instance in the data.

An evoked response potential (ERP) is the observed electrical activity that takes place

when a stimulus is presented to the human participant. Electrophysiology is a study of the electrical properties of biological cells, and typically refers to properties of single or small groups of cells. The traditional analysis of ERP involves averaging trials of ERPs under different experimental conditions and comparing the relative differences.

In this work, we refer to anatomical locations within the brain; specifically, the inferior-frontal and parietal regions. Within the parietal lobe, the intraparietal sulcus (IPS) separates the superior and inferior parietal gyri.

1.6 Feature Selection

Feature selection represents an subfield of machine learning research that focuses on finding relevant features in data. Typically this data is high dimensional, where the number of features of the data is greater than the number of instances in a given dataset. The goal is to find a minimal set of features that optimally predict class labels. This lower dimensional representation is often more interpretable than classification rules that use the complete feature set. This problem has been looked at from a number of perspectives, from nonlinear dimensionality reduction methods to feature discovery to feature selection. Feature selection has also been addressed from the perspective of signal segmentation methods. These approaches have focused on filter design techniques to factor noise signals as well as target signals. In this work, we explore the application of dimensionality reduction and sparse regularization to independent components analysis. The hypothesis is that in noisy and overcomplete situations, these methods will produce features that are more interpretable and that improve classification accuracy. The long-term aims of this work are to show the formal conditions under which these methods work. Another long-term aim is to put these methods in the hands of practitioners; we also demonstrate how slow nonlinear artifacts may be removed from EEG data using a spline regression strategy in terms that practitioners can use.

1.7 Psychophysiology: Mental Rotation

Mental rotation is one mental task among many that are used as a means of indicating intent in brain-machine interfaces. From a psychophysiology perspective, mental rotation is

a well studied phenomena that is at the heart of a number of debates in the psychophysiology community such as its role in the functional termination of the dorsal visual pathway. Today, a number of research directions in Electrophysiology and psychophysiology are converging on identifying functional neural correlates of mental rotation phenomena. Can we recover this structure in ERP functional data and what are the implications for application of mental rotation in brain-machine interfaces? We propose that orientation selective networks in the parietal lobe produce detectable activation differences in ERP data. We approach mental rotation from the perspective of evoking detectable potential differences for the purpose of control. The hypothesis is that mental rotation ERP is differentiable from language production ERP and that we can detect orientation selective activity in mental rotation ERP.

1.8 Symmetry perception

Symmetry perception is a mental task that lies between high and low level visual processing. Some research suggests that symmetry perception relies on top-down information as much or more than bottom up information. How does the visual system represent and integrate features of visual images to estimate symmetry in the early visual system? We propose an ERP based study aimed at measuring early ERP differences of manipulations of approximate planar symmetry. The results have implications for models of symmetry processing that rely on top-down information. We hypothesize that early ERP differences correlate with judgment (*i.e.* perceived symmetry).

1.9 Brain-machine interface

We hypothesize that sparse feature selection methods recover ERP segments relevant for classifying in a single-trial setting. The implication is that feature selection is an essential step in achieving reliable single trial classification of ERP data. Further sparsity is a general property of ERP signals with high information content; sparse regularization assists independent components analysis methods in classifying ERP signals.

1.10 Contributions

This dissertation investigates the use of sparsity in feature selection and makes three main contributions to the fields of machine learning, psychophysiology, and brain-machine interface research. We explore dimensionality reduction methods as a preprocessing step for ICA in fMRI data. We explore embedded sparse regularization strategies for ICA and present rICAs, a ICA method based that uses sparse regularization. We explore ERP differences in mental rotation as a candidate mental task for BMIs. We explore ERP differences in perception of approximate symmetry. We demonstrate the relative accuracy of rICAs for single-trial classification of ERP signals using the data from these studies. We also present a method of slow, nonlinear artifact removal.

1.11 Background

From a machine learning perspective, we are interested in classification over multi-dimensional data where we are able to obtain labeled data. Within-subject ERPs are generally the same for perceptual tasks; there are few learned effects that alter these signals. We pose ERP classification as a supervised learning problem, where we are able to gather labeled data of each respective class and learn a discrimination function for classifying future unseen instances of data. The objective is to find a minimal number of features while simultaneously minimizing classification error. For BMIs, there are a predetermined number of target classes we are interested in, further supporting a supervised learning approach.

In many supervised learning problems, we often want to select the features of the data that are most relevant (highly correlated, and discriminant) to a classification label. Often, many features are distractors, in that they do not contribute to the classification labels of interest. How can we find features within a dataset that are most relevant to classification? Dimensionality reduction problems aim to find lower dimensional representations of data using the hypothesis that the latent least-error classification function is usually of lower dimension than the data.

In an unsupervised setting, we are interested in discovering features and functions of features most relevant to classification, without the benefit of labeled data. For a large class

of these problems, statistically independent features are desirable. These features often are called the latent causes of the data that give rise to the observations, and classification over these features often provides additional interpretation of data. Methods for finding independent components are often brittle, often requiring a priori knowledge of the number of sources (independent features) and the sensors to be relatively noise-free. Embedded methods of regularization in these methods can address some of these deficiencies while making few other assumptions on the underlying causes.

Unsupervised feature selection has particular applicability in brain-machine interfaces and single-trial classification tasks. Precise inference using noisy sensor arrays like EEG require precise target signal segmentation after artifact removal. For target signals to be interpretable, signals should not only be smooth for optimization, they should also be sparse in activation.

1.11.1 Brain-machine interfaces

Supervised learning approaches have been applied to BMI data previously (91; 10; 28). Unsupervised methods are also typically used in BMI work as well as general ERP analysis. rICAs represents a novel application of sparse regularization to this data. The objective of this collective work is to produce methods capable of single-trial classification in ERP data.

Motivation for BMI research comes from several populations of human subjects with specific interface needs. An important population motivating research in assistive technology are locked-in subjects. Subjects experiencing traumatic brain injury (TBI) or debilitating diseases such as amyotrophic lateral scelorsis (ALS) experience a *locked-in* syndrome; where few channels are available for interaction with the environment, but cognitive ability remains largely functional. A basic goal of research in this area is restoring subjects' ability to interact with the environment; a primary method utilizes patterns of activation in the brain. These patterns can be used for control; the implication being these patterns are reliably detectable with high accuracy and require minimal cognitive effort to generate (117; 34; 64; 66; 82; 90).

Portable sensor arrays such as electroencephalography (EEG) are promising for applying neural signals to near real-time control tasks in general-use situations, because they provide minimally invasive sensing arrays for observing neural activity with good temporal resolution. While available for some time (55), EEG sensor arrays have typically been reserved for clinical use with limited application outside of select subject populations in part due to limited inference capabilities using noisy data (63). General use BMI is an emerging technology that offers a novel method of human-computer interaction. As electrode sensor technology advances usability of an EEG based interface becomes more general: fewer constraints of array application and improved models of evoked cortical activity combine to produce interfaces suitable for general use.

1.11.2 Artifact removal and BMI

Artifact removal describes the process of identifying artifact signals and trends and separating them from target signal. The goal is to separate artifacts from target signals that preserves the information contained in the target signals. Typical approaches to artifact removal include filter design or defining parameters of signal processing filters that minimize information loss in target signals. The basic properties of filter design do not filter trends whose frequency characteristics are similar to target signals. Component analysis methods are also used to remove artifact components. The goals of the method defined here are to remove low-frequency nonlinear artifacts while making the method accessible to practitioners.

1.11.3 Pyschophysiology perspective

From a Psychophysiological perspective, we are interested in finding mental tasks that produce differentiable ERPs using the same stimulus or evoking activity in the same region of the brain. These tasks are therefore controllable in contrast to P300 or steady state responses. Mental rotation is a well studied task requiring aspects of the low level visual system as well as high level representation information. Symmetry assessment is also such a task; both are evoked with visual stimuli.

1.12 Thesis statement

The objectives of the work presented here support the following claim:

Sparse regularization improves components analysis in noisy, overcomplete environments; a psychophysiological analysis of mental rotation shows its applicability to BMIs. The combination of these approaches enables more robust BMI.

1.13 Long-term goals

The long term goals of this work are aimed at formalizing sparse regularization performance. Formal descriptions of how sparse regularization accounts for noise are needed to show its broad applicability to other machine learning problems. Further study and exploration of mental rotation in terms of orientation selective activation is needed to expand its use in BMIs.

1.14 Committee members

Charles Isbell, Melody Jackson, Paul Corballis, Alex Gray, Thad Starner

CHAPTER II

DIMENSIONALITY REDUCTION AND ICA IN FMRI DATA ANALYSIS

2.1 Introduction

We present two methods in source separation techniques aimed at improving interpretability of data in two different types of analysis. First, in fMRI data analysis, region-of-interest selection is a common practice; selecting anatomical regions at or after acquisition time makes analysis tractable. ROI selection is susceptible to type II errors and can affect results. We use manifold learning techniques to reduce the dimensionality of fMRI volume data. We compare our analysis to previous analysis of volume data, demonstrating improved feature discovery. Second, a primary analysis tool of ERP data is independent components analysis. Finding the underlying causes in ERP data gives researchers insight into finding response differences between experimental and control data. Unfortunately, ICA methods make an assumption of completeness: that the number of sources is equal to the number of sensors. In ERP data, this is rarely the case. Often sensor arrays are sparse and subject to noise from the environment as well as the subject. Further, in dense array conditions, it can be difficult to interpret the number of sources that are recovered. Being able to recover sources that are less noisy and more easily interpreted enables better analysis of ERP data. Our objective is to select components and features that best represent target signals while maximizing interpretability of components. We propose a regularization method that optimizes this trade-off using ℓ_1 norm and approximations of ℓ_0 norm. We explore $\ell_n : 0 < n < 1$ norms in this context. We show the broad applicability of these methods to other source separation problems. This represents an analysis tool for ICA users, particularly with ERP data. This also represents an important contribution to machine learning research, combining regularization and source separation.

2.2 Independent Components Analysis

We first give the general ICA problem setup described as a linear transformation on the data, X. An ICA algorithm learns an unmixing matrix: A^{-1} that maximizes statistical independence of the components resulting matrix: $A^{-1}X = S$. The unmixing matrix is invertible such that $AS = X$. Several classes of strategies have been proposed for unmixing matrix discovery, each with several implementations of the general strategy. Here we examine three implementations focusing on entropy and negentropy optimization strategies. In these cases, we will apply sparse regularization within the defined optimization strategy for each method respectively. In general, we will refer to the regularization function within each optimization as $s(X)$.

Independent components analysis can be used in a feature selection setting. Today, many methods exist to generate components whose pairwise mutual information is minimal. There are many proven methods for generating this condition. Perhaps the most used of these methods is the infomax method, and the FastICA methods. In the case of infomax ICA, the optimization is to maximize the mutual information between data and the monotonic transform of the data. Using a monotonic transform, the optimization can be put in terms of the data itself, subject to scaling. At a high level, infomax ICA finds a matrix that aligns the high density parts of the data distribution with high density parts of the transform distribution.

ICA is a method for discovering a transform over the data that generates statistically independent components which are linearly mixed to produce the original data. The transform is invertible, so that the data may be linearly unmixed back into the components. The intuition is that statistically independence is an important property of underlying causes of the data:

$$\max I(Y, X) \tag{1}$$

$$y = g(x) \tag{2}$$

$$f_y(y) = \frac{f_x(x)}{|\frac{\partial y}{\partial x}|} \tag{3}$$

$$\tag{4}$$

where $I(Y, X)$ is the mutual information function between the output Y and input X matrices, $g(x)$ is the first derivative of the contrast function, $f_n(n)$ is the probability density of the inputs and outputs (x and y respectively). The intuition for Infomax ICA is given by the following principles:

- Reducing redundancy between output sources is desirable, and maximally statistically independent sources are important to subsequent analysis.
- The objective of infomax ICA is to minimize pairwise mutual information between sources.
- Minimizing pairwise mutual information between sources is equivalent to maximizing the mutual information between corresponding inputs and outputs
- Maximizing mutual information between corresponding inputs and outputs is equivalent to maximizing the Shannon entropy of the transformed data, using a transform that is monotonic
- The gradient to this optimization is with respect to the weight matrix w, where we want to use w to center the mass of the data distribution on the center of the mass of the transform distribution.

The assumption that monotonic functions are needed to generate the target distributions may be generalized to one-to-one functions. As long as each point has a unique representation in the range of the transforming function, there exists a matrix that can map high density points in the original representation to high density points in the target distribution.

2.3 Dimensionality reduction for source separation

Functional magnetic resonance imaging (fMRI) captures neural activation patterns by measuring the hemodynamic response in cranial tissue through sampling discrete regions of the brain, referred to as voxels (33). Each voxel represents the aggregate hemodynamic response of a region of neurons. Locating significant differences between active and non-active voxels is challenging because of the inherent latencies and artifacts in fMRI signal acquisition (60). Furthermore, the hemodynamic activation level of neighboring voxels influences voxel activation, producing less accurate spatial activation maps.

Behavioral experiments using fMRI typically evoke activation in a hypothesized *region of interest* (ROI) in the brain. The ROI represents an anatomical region of the brain believed to be where functional processing of a specific behavioral task occurs. Experimental trials in these designs use a behavioral task meant to evoke activation in the ROI. Control trials do not evoke ROI activation.

Traditional analysis methods such as statistical parametric mapping (SPM) use statistical tests to demonstrate significant differences between control and experimental voxel time courses (43). The objective of component analysis methods is to recover components whose time course activation correlates with the task-based reference function: $\text{argmax}_{a \in \{A\}} \rho(r, a)$, where r is the reference activation time course that represents the ideal activation during the trial, and a is the component activation time course. Typical separation methods such as independent components analysis (ICA) do not account for delayed composition effects.

We frame the problem of the combined latencies of the hemodynamic response and the signal acquisition process as a convolution of the hemodynamic response functions of spatially independent components. In order to address these confounding spatial and temporal influences, we first use nonlinear manifold learning to constrain source separation and to remove voxels which do not help distinguish between task and non-task activation. We generate a frequency space representation of the reduced features for convolutive source separation. Finally, the resulting components are visualized by plotting the power spectra of both sources and activation time courses.

Furthermore, canonical ICA methods have been shown to work for simple block experimental designs where the ROI is understood a priori (79). An important property of automated methods for activation recovery is the ROI selection. Selecting the ROI should be without a priori anatomical or activation area information. Furthermore, the ROI selection should be unconstrained to limit type II errors. The intuition is that by allowing more voxels to be analyzed in a similar amount of time, we reduce that chance that ROI selection eliminates relevant voxels from analysis. More importantly, our method addresses this by allowing for larger ROI selection, without specific a priori anatomical knowledge of the ROI.

Our approach to fMRI analysis seeks to combine the strengths of manifold learning, convolution in frequency space, and complex ICA in order to improve the accuracy of recovered brain activity components. Using ICA in the frequency domain allows us to treat convolution of components as a product, which in turn allows a computationally feasible algorithm to solve the convolutive blind source separation problem. Using this version of the source separation problem is important since voxels near each other in the brain may exhibit delayed influences during recording. Using a convolutive model rather than an instantaneous mixing model provides the ability to capture this influence and properly separate the components.

We use a set of existing dimensionality reduction techniques to generate lower dimensional representations of imaging data: diffusion maps (Diff Map) (19), isomap (104), Local Linear Embedding (LLE) (97; 38), and LaPlacian Eigenmaps (LE) (8).

2.4 Dataset 1

To evaluate our method, we begin with a simple example: consider an fMRI scan sequence of a single subject performing a repetitive right- or left-hand finger movement task (54). The objective is to find task related activated components of hand movements in the ROI. For the ROI, we selected a window of voxels in a region based on correlation values to the reference function using time domain ICA. In the motor task, 80 volumes were sampled at a constant rate for each task: left-hand/right-hand finger move. We defined a ROI in

slices 13,14,15,16, loosely defined around the temporal area of the motor cortex. Scans of left hand tasks are concatenated to scans of right hand tasks, 160 scans total. Given this organization, the reference activation function for left hand tasks is defined as a delta function: $\delta(x \geq 80)$.

First, we want to test how manifold learning techniques assist in time domain separation. In this case, we compare the correlation of component activations recovered by ICA to reference function activation. We compare the best correlation values generated using ICA alone as well as with the various manifold learning techniques. These are all performed using the time domain data (see Table 1). The manifold learning methods do not recover correlated activation of components as well as using ICA alone in this case.

Table 1: Comparison of correlation values to reference function using manifold learning in time domain.

Method	Max ρ	p-value
Diff Map	0.1407	0.1
Isomap	0.3470	0.001
LLE	0.2052	0.01
LE	0.2236	0.005
ICA	0.7395	0.0001

For the STFT, we use a parametrization for each dataset. In the case of the left/right dataset, the window size is the ratio of the hemodynamic response latency to volume acquisition latency. We consider the measured values of voxels v_i through time $t_{i\in\{1\ldots\tau\}}$. Each STFT generates frequency vectors for each window. We group frequency vectors from each STFT and apply component analysis to the resulting matrices. Computing the inverse transform of the component produces a time domain representation of the signal. However, due to the window overlap in the STFT, this time scale is not appropriate for comparison in the original observation space.

We compare the performance in the left/right task between the various manifold learning algorithms and complex ICA in the frequency domain without manifold learning (see Figure 1). To compare methods, we use the minimum distance of component activation to reference function activation in each frequency bin. In this case, manifold learning using

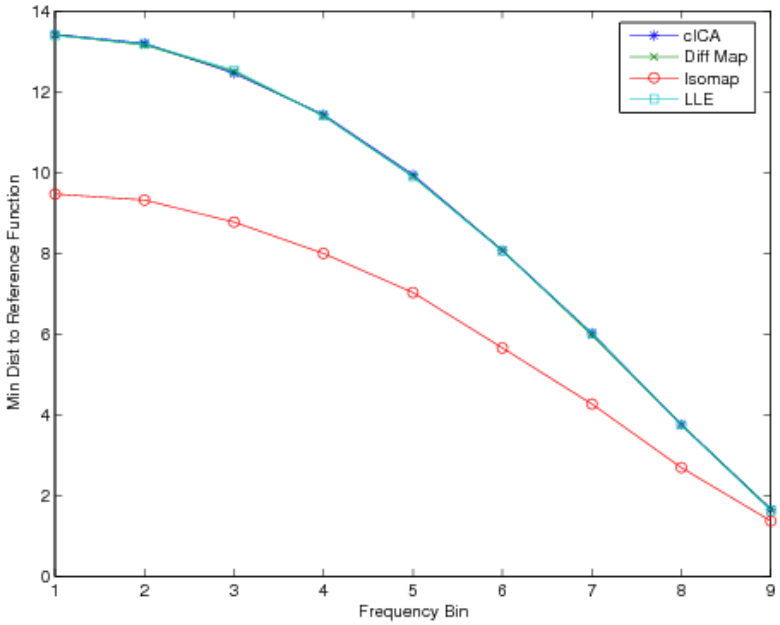

Figure 1: Comparison of minimum distances to reference function between manifold learning method preprocessing and complex ICA. Minimum distance for each method in each STFT frequency bin.

diffusion maps and local linear embedding perform slightly better than complex ICA alone.

2.5 Dataset 2

Postle *et al.* (93) measured activation of five participants in four behavioral tasks: forward memory, manipulate memory, guided saccade, and a free saccade task. Subjects completed 96 trials: 8 blocks of 12 trials each. Within each block, subjects received an equal number of task trials, in random order. Subjects were presented with a static arrangement of squares on a screen. Signals were acquired using a GE 1.5T scanner with 3.75mm^2 in-plane resolution and 5mm inter-slice distance. Volumes were 21 slices, and volume acquisition time was 2s; 17 volumes were acquired per trial. Inter-trial time was 7s. By comparing voxel activation values in each experimental task in the ROI, Postle *et al.* showed no significant difference in voxel activations between control and experimental tasks.

Figure 2 shows the trial sequence. First, subjects are told what the trial task will be: "memory," "no memory," or "free eye movements." Following an interstimulus interval

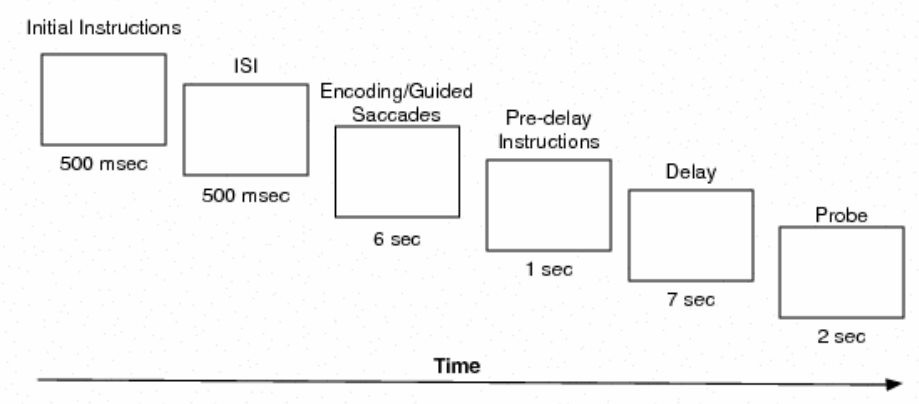

Figure 2: Trial event sequence (93). Initial instructions indicate what the memory task will be. After ISI, a sequence of highlighted boxes (see Figure 3) or fixation points appear. Pre-delay instructions indicate whether the memory task is "forward," "down to up," or "fixate." After the delay, the probe is shown.

(ISI) of 500ms, subjects receive the sequence of highlighted squares (Figure 3) followed by further instructions: "forward," "down-to-up," or "fixate." After the return to baseline delay, subjects receive the probe: a highlighted square in the sequence.

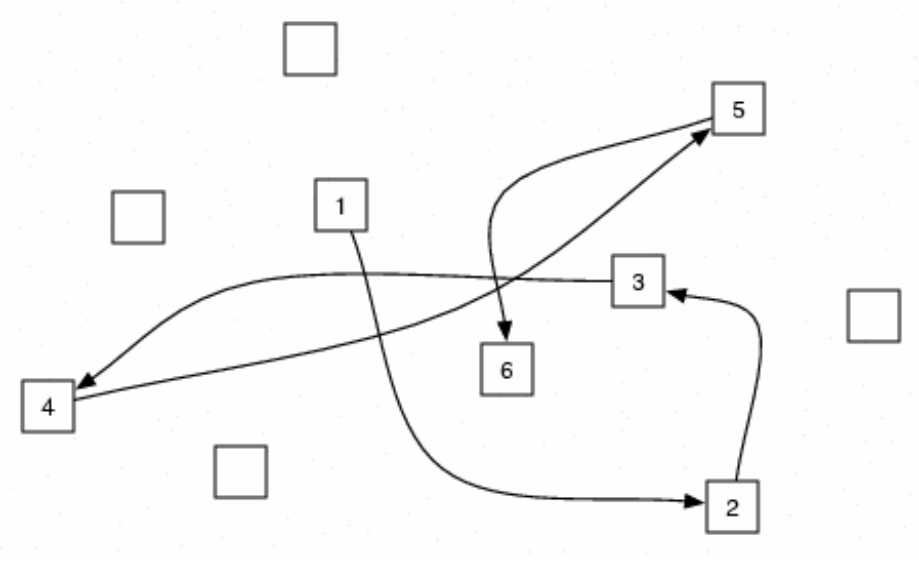

Figure 3: Memory task stimulus. A fixed number of squares are oriented on a screen. During memory tasks, a sequence of the squares are highlighted in a random order. An example highlight sequence for memory is shown.

Behavioral Tasks. During forward memory, manipulate memory, and guided saccade tasks, a sequence of squares was highlighted followed by a delay and then a task prompt (see Figures 2&3). In forward memory tasks, subjects were presented a sequence of highlighted

squares. Then, given one of highlighted squares, subjects were asked to recreate the sequence from that point on. In the manipulate memory task, subjects were asked to reorder the highlighted sequence of squares from bottom to top, so that the lowest highlighted square should be first in the sequence and the highest should be last. In the guided saccade task, subjects were asked to simply follow another highlighted sequence on the screen. In the free saccade task, subjects were not shown a highlighted sequence, and were asked to simply saccade left and right repeatedly.

In these experiments, we consider a ROI based on the reported areas in each subject. We constrain the ROI to be even smaller. In this experiment, we use the manipulate memory task as the experimental task alone and generate the reference function for each subject.

Time Domain Experiment. We apply the method to time domain signals, as in the left/right task. In this case, dimensionality reduction methods produce signals that do not compare on the time axis. In this case, we compare the correlation of the power spectra from activation time courses to the reference power spectrum. Here we compare the first 50 frequency values, accounting for over 99% of the frequency content in the reference signal. ICA generated components are well correlated across subjects. However, local linear embedding appears to outperform ICA in subjects H and S (see Table 2).

Table 2: Time domain comparison using Postle *et al.* dataset. Correlation of power spectra for activation time courses generated for each subject using ICA and the various dimensionality reduction methods: ICA (ICA alone), Isomap (Isomap and ICA), LE (Laplacian eigenmap and ICA), and LLE (Local linear embedding and ICA).

Subject	ICA	Isomap	LE	LLE
H	0.7771	0.6944	0.8600	0.8774
K	0.9412	0.7288	0.8229	0.7897
S	0.8423	0.7319	n/a	0.8719
T	0.8903	0.7657	0.8274	0.8094
W	0.9262	0.7156	0.8268	0.8711

Frequency Domain Experiment. We apply the method to the frequency domain signals using the same comparison method used in the left/right dataset. In this case, dimensionality reduction methods outperform ICA alone for most subjects (see Table 3. For subject H, Isomap appears to recover sources whose activation better matches the reference function. For subject K, ICA alone appears to outperform the manifold learning

methods. For subjects S,T, and W, manifold learning appears to generate better source separation.

In our proposed method, the motivation for using manifold learning as a pre-processing step to convolutive source separation is dimensionality reduction. We are interested in identifying voxels that contain the activation information for the ROI. Furthermore, the frequency space representation of voxels generates a much higher dimensionality; reducing the dimensionality is critical to component analysis. The computational cost of filtering unneeded dimensions at component analysis time is far greater than at manifold learning time.

A typical pre-processing step when using time domain ICA is to perform PCA to find a projection that whitens the data. In the time domain PCA decorrelates the data, making the source separation task return better results. The idea in using manifold learning to reduce dimensionality is that we identify the voxels in the ROI that contain the most information about the activation sequence of the area. Performing dimensionality reduction using methods such as isomap produces lower dimensional representations of the data; if these representations are accurate, then the components produced using ICA should be accurate as well. The computational benefit at source separation time allows for a larger number of variables (voxels) to be considered. In situations where the dimensionality is reduced to spaces allowing for visualization, we expect the resulting components to also be more interpretable.

Table 3: Comparison of minimum distances to reference function activation between manifold learning methods in combination with complex ICA and complex ICA alone. For each bin (columns), the minimum distance for each method is shown (*i.e.* the distance of the best matching components in each frequency bin).

	1	2	3	4	5	6	7	8	9
Subject H									
ICA	30.9077	29.5631	28.0292	25.5206	22.1484	18.0443	13.3546	8.1448	2.8922
Isomap	29.9395	29.4552	27.8957	25.4514	22.1252	18.0155	13.3221	8.1521	2.8443
Diff Map	30.2166	29.6940	28.1611	25.6646	22.3125	18.1816	13.4606	8.2916	2.9831
LLE	30.2003	29.6894	28.1652	25.6854	22.3254	18.1924	13.4682	8.2789	2.9805
Subject K									
ICA	30.1220	29.5681	28.0250	25.5203	22.1783	18.0569	13.3136	8.2129	2.9432
Isomap	30.1864	29.6640	28.1538	25.6675	22.2939	18.1749	13.4683	8.2771	2.9618
Diff Map	30.2093	29.7037	28.1519	25.6816	22.3158	18.2034	13.4654	8.2952	2.9904
LLE	30.2080	29.6818	28.1548	25.6716	22.3262	18.1962	13.4695	8.2875	2.9790
Subject S									
ICA	30.2240	29.7156	28.1836	25.6819	22.3365	18.2140	13.4956	8.3269	3.0025
Isomap	30.2044	29.7038	28.1697	25.6823	22.3215	18.2035	13.4576	8.2912	2.9832
Diff Map	30.2066	29.6898	28.1527	25.6871	22.3178	18.1870	13.4652	8.2941	2.9765
LLE	30.1965	29.6953	28.1583	25.6731	22.3160	18.1937	13.4655	8.2990	2.9525
Subject T									
ICA	30.2336	29.7114	28.1868	25.6938	22.3342	18.2210	13.4869	8.3183	3.0027
Isomap	30.2077	29.6922	28.1493	25.6869	22.3167	18.1991	13.4576	8.2959	2.9792
Diff Map	30.2115	29.6900	28.1641	25.6714	22.3120	18.1877	13.4735	8.2751	2.9868
LLE	30.2100	29.6754	28.1444	25.6737	22.3142	18.1942	13.4569	8.2977	2.9799
Subject W									
ICA	30.2307	29.7087	28.1928	25.7006	22.3436	18.2214	13.4755	8.3068	2.9984
Isomap	30.1833	29.6769	28.1525	25.6727	22.3140	18.1835	13.4538	8.2765	2.9460
Diff Map	30.2106	29.6915	28.1508	25.6780	22.3146	18.1972	13.4617	8.2923	2.9748
LLE	30.2044	29.6896	28.1588	25.6591	22.3199	18.1946	13.4617	8.3007	2.9805

CHAPTER III

SPARSE REGULARIZATION IN ICA

3.1 Introduction

Independent components analysis (ICA) methods produce transformations for maximizing statistical independence of data, typically features. The motivation is to generate features that represent the underlying causes of observations, increasing the interpretability of the original data. ICA as a transformation method over features may also serve dimensionality reduction motivations, where the transformation is a lower dimensional representation of data. In many practical instances, blind source separation methods limit interpretability of recovered sources in overcomplete cases where the number of sources is not equal to the number of sensors. Indeed, it is rarely the case in data driven analysis that the number of underlying sources is known a priori or that the number of sensors may be manipulated to accommodate the estimated number of sources. The result is that while ICA methods still recover maximally independent sources, the method is forced to produce mixtures of the remaining sources. Particularly in the case of noise, residual source mixtures limit interpretability. The objective of this work is to produce regularized blind source separation methods that produce better source estimates in noisy and overcomplete cases.

We first give the general ICA problem setup described as a linear transformation on the data, X. An ICA algorithm learns an unmixing matrix: A^{-1} that maximizes statistical independence of the components resulting matrix: $A^{-1}X = S$. The unmixing matrix is invertible such that $AS = X$. Several classes of strategies have been proposed for unmixing matrix discovery, each with several implementations of the general strategy. Here we examine three implementations focusing on entropy and negentropy optimization strategies. In these cases, we will apply sparse regularization within the defined optimization strategy for each method respectively. In general, we will refer to the regularization function within each optimization as $s(X)$.

3.2 Regularization in feature selection

In a supervised learning setting for linear models, a set of instances $x_i \in \Re^p, i = 1 \ldots n$ are given with a set of corresponding labels $y_i, i = 1 \ldots m$. The objective is to learn a set of weights β that minimize loss between predicted labels and the true labels for previously unseen data. A competing objective in learning optimal weights is generating a minimal number of non-zero weights. Ideally, a linear model would assign non-zero weight only to those variables most relevant to the model. Feature selection methods are concerned with the tradeoff between the size of the model and minimizing the loss of the model.

Feature selection methods address several important issues facing many learning problems: the curse of dimensionality, generalization of the learner, model interpretability, and scalability. Clearly, reducing the size of the model while maintaining classification accuracy selects features most relevant to classification with respect to the model class. Fewer features in the model improve the interpretability of the model. Models with fewer variables will scale as the amount of data increases.

In a regularization setting, the objective of optimizing a loss function of the model is pitted against the objective of optimizing a penalty term. The penalty term typically is a norm of the weight vector (ℓ_1 and ℓ_2 respectively) $\lambda \sum |\beta|, \lambda \sum \beta^2$

$\lambda \in (0, 1)$ is a parameter that represents the relative weighting between the loss term and the penalty term. These norms are typically used in part for their ability to generate feature weights that may be used to select features, but also for their tractability. In contrast, the ℓ_0 norm seems a better penalty term for selecting sparse models:

$$\lambda \sum_i x, x = \begin{cases} 1 & \beta_i \neq 0 \\ 0 & \beta_i = 0 \end{cases} \tag{5}$$

This norm directly penalizes non-zero weights, and therefore should produce maximally sparse models in a regularization setting. However, feature selection using ℓ_0 regularization is not tractable due to the fact that the ℓ_0 function is not differentiable. While ℓ_1 performs sparse feature selection, true ℓ_0 feature selection may be preferable in many situations, particularly where $n \ll p$, that is the number of features is greater than the number of

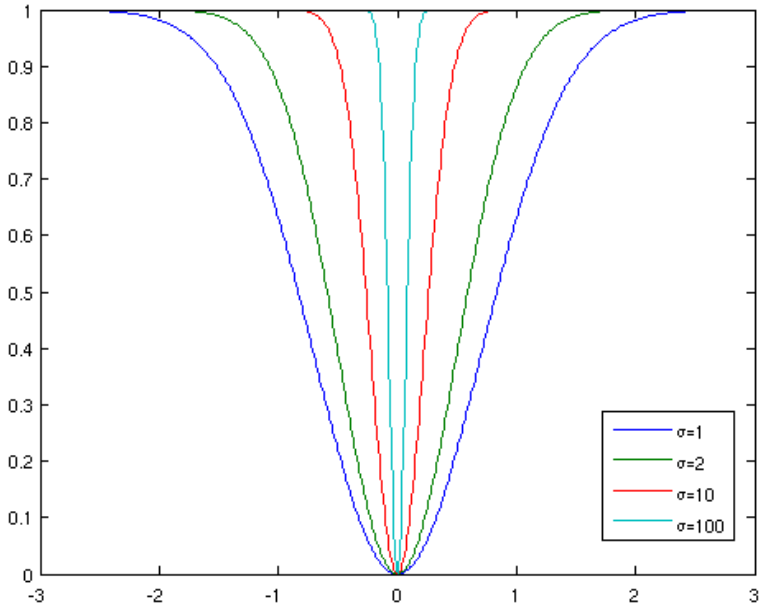

Figure 4: $\hat{\ell_0}$ behavior about zero. $\lim_{\sigma\to\infty} \int \ell_0 - \int \hat{\ell_0} = 0$

instances. This condition is typical in many genres of data, particularly in medical imaging settings where observation sets are expensive to gather, but each observation set is high resolution for observing certain phenomena.

3.3 ℓ_0 approximator

We focus on a particular form of the common ℓ_0 approximator: $1 - e^{-\sigma\beta^2}$. This form has a number of properties that make it useful to regularization within many forms of ICA algorithms. First, the approximator is differentiable, making it amenable to unconstrained optimization methods. Figure 4 shows $\hat{\ell_0}$ for increasing values of σ; the function gets arbitrarily close to ℓ_0 as σ increases. Finally, as a regularization term, $\hat{\ell_0}$ should provide more aggressive feature selection.

3.4 Previous work

Approximations for the ℓ_0 norm have been proposed previously. Bradley and Mangasarian (12) suggest a similar equation for ℓ_0 approximation in regularization of support vector machines that is also used by Weston et al. (116):$1 - e^{-\alpha|v|}$, $\alpha > 0$. Using this norm

approximation, a linear programming solution is used for feature selection. Bradley and Mangasarian (12) use the results of the optimization in a support vector machine to regularize the max margin function. The objective is to increase between class distance to minimize misclassification of test instances.

Earlier work on ℓ_1 and ℓ_2 regularization laid the foundation for other regularization methods. Previous solutions have focused on exact solutions using regularization terms where closed form solutions exist in the ℓ_2 case (105; 122). In the ℓ_1 case, linear programming solutions have been successful. Least angle regression is one such method for solving ℓ_1 (lasso) based regularization for linear models (36). More recent work has focused on solving specific problems with the lasso solution. For instance, in cases where there are more features than instances ($p > n$), lasso selects at most n features, even when a more complex model is needed. Lasso regression also requires a fixed value parameter representing the bound on the ℓ_1 norm. These methods also focus on discovery of higher order relationships in data. Correlation between features adversely affects lasso performance, where ℓ_2 (ridge) regression has been shown to dominate lasso performance. The elastic net is one proposed method to address these problems (122). It uses a combination of ℓ_1 and ℓ_2 norms for feature selection: $\hat{\beta} = \text{argmin}_\beta |y - X\beta|^2, \text{where } (1-\alpha)|\beta|_1 + \alpha|\beta|^2 \leq t$. Where α is a tradeoff parameter between the two norm terms, and t is a tuning parameter for the overall regression. Informally, this penalty term combines the strengths of both ℓ_1 and ℓ_2 norm penalties. Zou and Hastie also present LARS-EN, a Lars analog algorithm for the elastic net penalty (122).

Previous work has also addressed problems of Gaussian distributed signals in source separation tasks. These methods have used properties of the target separation sources or noise models to factor confounding sources (17; 113; 14). Here, we propose appealing to sparsity as an additional objective to independence in optimization.

Sparse component discovery has recently been observed in image data (5; 13). Regularization using sparse penalty functions has also been applied to principal components analysis (26; 123). The motivation for sparse regularization in PCA is to improve interpretability of principal components by enforce sparse contributions of sources to data

reconstruction.

Finally, research in support vector machines (SVM) has increased the power and popularity of linear models in classification tasks. Once thought to be less favorable to nonlinear model learning, linear models are once again popular due to the kernel trick. Methods that use the kernel trick use functions of higher dimensional spaces where linear models separate data in place of more expensive dot products between data points.

3.5 Infomax ICA

We consider the infomax approach to blind source separation, adding the regularization term to the initial objective: $\text{argmax}_w\, I(Y,X) - s(Y)$. The infomax objective maximizes information flow between input and output. Mutual information is based on Shannon differential entropy: $H(\mathbf{y}) = -\int f(\mathbf{y}) \log f(\mathbf{y}) d\mathbf{y}$. The derivation of the gradient rule proceeds as proposed by Bell and Sejnowski in (9), dropping $H(Y|X)$ and $\ln f_x(X)$ as the derivative is not affected by these terms. In this case, the motivation for the choice of the regularization function becomes clear: we are able to use the existing gradient method with the additional regularization term to achieve sparse components. We develop the derivation for the regularization term using the strategy presented by Bell and Sejnowski in (9).

3.5.1 Single input, output

We introduce the regularization term to the single valued input and output example used by Bell and Sejnowski in (9). Here, the gradient descent optimization for a single input, with the regularization term added is: $H(y) = -\mathbb{E}[\ln |\frac{\partial y}{\partial x}|] - \lambda e^{-\sigma y^{-2}}$. We derive the second term on the right hand side:

$$f = -\lambda e^{-\sigma[1+e^{-wx-w_0}]^{-2}} \tag{6}$$

$$\frac{df}{dw} = \frac{df}{da}\frac{da}{dw}, a = -\sigma(1+e^{-wx-w_0})^{-2} \tag{7}$$

$$\frac{df}{da} = e^a \tag{8}$$

$$\frac{da}{dw} = \frac{da}{db}\frac{db}{dw}, b = 1+e^{-wx-w_0} \tag{9}$$

$$\frac{da}{db} = 2\sigma b^{-}3 \tag{10}$$

$$\frac{db}{dw} = \frac{db}{dc}\frac{dc}{dw}, c = -wx - w_0 \tag{11}$$

$$\frac{db}{dc} = e^c, \frac{dc}{dw} = -x \tag{12}$$

$$\frac{df}{dw} = -\lambda e^{-\sigma(1+e^{-u})^{-2}} \cdot 2\sigma(1+e^{-u})^{-3} \cdot e^{-u} \cdot -x \tag{13}$$

3.5.2 NxN network

In the case of multiple inputs, the objective is given as:

$$\frac{\partial H}{\partial w} = \frac{\partial}{\partial w}(\ln|\frac{\partial Y}{\partial X}|) - \frac{\partial}{\partial w}\lambda(1+e^{-\sigma Y^T Y}) \tag{14}$$

The derivation proceeds as in (9) with the additional regularization term to yield the update rule:

$$\frac{1}{w} + (x(1-2y)) - 2\lambda y e^{-\sigma y^T y} \tag{15}$$

3.5.3 Experiments

The experiments are aimed at testing the relative performance gains of sparse regularization in source separation. In particular, the prediction is that sparse regularization will improve separation of target signals in square and overcomplete noisy situations. We test this prediction using random sampled sources from continuous kurtotic distributions (Cauchy, exponential, f-distribution, gamma, Laplace, and Rayleigh). We generate noisy sources by sampling from parameterized Gaussian distributions. The parametrizations of these distributions were controlled so that a non-trivial amount of overlap occurred for each case. In both experiments, a random mixture matrix was generated sampling from the unit Gaussian. The experiments test performance differences in two situations: a square case (*i.e.*

$n = p$) where one of the objective sources is Gaussian distributed and an overcomplete case where one source is Gaussian distributed. In both experiments, we test relative performance against increasing set size. For each sample size, we generated 200 cases and compute the sample mean performance. In both experiments, infomax ICA methods were applied where the learning rate initial value was 10^{-3}. Ten epochs were run before reducing the learning rate by an order of magnitude until the learning rate was 10^{-9}. In general, default parameters were used to test performance of each ICA method.

The first experiment tests separation in a square case where one source is Gaussian distributed and the other is one of the continuous kurtotic sources. We apply infomax and sparse regularization infomax to the source mixtures and measure the relative error in the recovered unmixing matrix using Amari error (17):

$$\frac{1}{2d(d-1)} \sum_{i}^{d} \left[\frac{\sum_{j}^{d} |b_{ij}|}{\max_j |b_{ij}|} - 1 \right] + \frac{1}{2d(d-1)} \sum_{j}^{d} \left[\frac{\sum_{i}^{d} |b_{ij}|}{\max_i |b_{ij}|} - 1 \right] \tag{16}$$

The metric provides a scale insensitive means to measure error in source estimates by measuring distance from each entry of the unmixing matrix to the inverse of the true matrix.

The second experiment tests overcomplete cases where one source is Gaussian distributed. Here, we generated two random samples from the continuous kurtotic distributions in the previous experiment, and the additional source was sampled from the unit Gaussian. A random mixing matrix was sampled from the unit Gaussian; two of the three mixtures were chosen and ICA methods were applied.

To assess performance in this case, we compare pairwise the randomly generated sources with the recovered sources. We use the sum of the minimum absolute value of the Euclidean inner product between pairs of sources to measure the relative divergence of recovered sources to ground truth. Sources were compared without replacement.

3.5.4 Results

Figure 5 shows the sample mean Amari errors for increasing samples sizes for each continuous, kurtotic distribution. In most cases, sparse regularization performed better or equal to infomax alone. In the gamma and Rayleigh distribution tests, performance was comparable

between the methods; particularly as source size increased. In a majority of these source sizes, sparse regularization produced lower Amari errors. In the F-distribution case, sparse regularization methods follow a clearer trend than in the other distribution cases. Here, sparse regularization methods perform comparable to infomax alone with infomax alone outperforming sparse regularization in only two cases. In the Laplace distribution test, sparse regularization methods produce lower Amari estimates than infomax where in each source size different sigma values affect performance.

For smaller source sample sizes, infomax ICA as a method generates sources with larger error. The intuition is that the gradient method requires more data to better approximate the distributions of the independent sources. As a result, infomax as a method is susceptible to outlier behavior, and variance as source size increases is generally attributed to outliers. For instance, in the Cauchy distribution test, outliers seem to have a larger effect on regularization methods in performance than in the other distributions.

Figure 6 shows the sample mean of the sum of the minimum Euclidean inner product for recovered sources versus source size for the distributions. In the Cauchy and Rayleigh distribution experiments, the trend of the sparse regularization methods is contrasted with the trend in the infomax case. The regularization methods follow a similar performance trend that differs from infomax alone performance. In these cases, in a majority of trials, sparse regularization infomax produced mixing matrix estimates with lower Amari error than infomax alone. For these distributions, different values of sigma yield similar results compared to infomax alone. In the case of gamma and Laplace distributions, a clear trend for sparse regularization methods does not seem to exist; different values of sigma affect the performance in different ways. Finally, sparse regularization shows improved source separation in the F-distribution experiment. The different sigma values tested show similar performance; all of them markedly improved over infomax alone.

3.6 RADICAL

While these results show improved performance for sparse regularization in the infomax case, we want to show a broader effect; that sparse regularization improves performance for

other ICA optimization strategies. In particular, RADICAL is a popular ICA method where we predict improved results using sparse regularization within each respective optimization strategy.

The RADICAL method of ICA maximizes an entropy measure based on order statistics of data (72). In the case of RADICAL, pairs of candidate components are considered, and an incremental search occurs over candidate rotation matrices to minimize the cost function based on its entropy estimate. The cost function allows us to add the regularization term directly without modification to the optimization function $\sum_i \log_2 x_i$.

$$\sum_i \log_2 x_i + s(x_i) \tag{17}$$

3.6.1 Experiments

The fact that we are able to apply functions to the RADICAL cost function without solving an optimization in RADICAL allows us to test other regularization terms, particularly ℓ_1 and ℓ_2. To test performance, we use the same experiment setup as in section 3.5.3, where we test noisy instances in square and overcomplete situations. A second set of experiments tests $\hat{\ell_0}$ performance for fixed values of sigma, using the same method as in section 3.5.3.

3.6.2 Experiment 1 results

First, we compare performance between different regularization terms in the square separation task. Figures 7 and 8 show the relative performance differences for ℓ_1,ℓ_2, and $\hat{\ell_0}$ regularization methods in RADICAL against RADICAL alone. In general, using a regularization term produces similar or better results in most cases of the square noise condition. In the Cauchy, F-distribution, gamma, and Laplace experiments performance is largely similar for the regularization methods and RADICAL alone. In the exponential case, the largest performance gains are seen using ℓ_1 and ℓ_2 regularization. With the exception of the largest sample size in the Rayleigh experiment, regularization methods performance comparable to RADICAL alone in Amari error.

Next, we compare performance in the overcomplete separation task. Again, in general,

adding regularization terms improves separation quality. In a majority of set sizes across all the distribution experiments, regularization methods produced lower sample mean euclidean product values. $\hat{\ell_0}$ regularization improved separation or performed comparable to RADICAL alone in the exponential, Laplace and Rayleigh experiments for a majority of sample sizes.

3.6.3 Experiment 2 results

We compare source recovery performance in the square and overcomplete cases for fixed sigma values in the $\hat{\ell_0}$ regularization term against RADICAL alone. In the Cauchy, exponential, F-distribution, and gamma tests, $\sigma = 1e-6$ performed comparable to or better than RADICAL alone. In general, sparse regularization improved separation results in the overcomplete case.

3.7 FastICA

FastICA is a popular method for source separation that uses negentropy in its optimization. Negentropy using Shannon differential entropy is defined as $J(\mathbf{y}) = H(\mathbf{y}_N) - H(\mathbf{y})$ where $H(\mathbf{y}_N)$ is the differential entropy of a Gaussian distributed source and mutual information using negentropy is defined as $I(\mathbf{y}) = J(\mathbf{y}) - \sum_i J(y_i))$. Using this definition, FastICA uses the difference of expectations of first and second order derivatives of a contrast function G (g and g' respectively) applied to data in a batch process (56). For a single input, the update rule for fastICA is given:

$$\mathbb{E}(wg(w^T x)) - \mathbb{E}(g'(w^T x))w \tag{18}$$

where each step of the batch process produces a new w estimate: $w_n = \frac{w_p}{||w_p||}$. Here, we add sparse regularization to the stepwise estimate function, and the algorithm proceeds as previously defined.

The experiments for fastICA versus sparse regularization fastICA were the same as those for infomax ICA (see Section 3.5.3). For fastICA, we used the default parameter settings for the experiments: using $G(u) = \frac{u^4}{4}$ as the root contrast function and the deflation optimization approach.

Again in general, $\hat{\ell_0}$ based regularization improves source recovery. In a majority of sample sizes, regularization methods produced unmixing matrices with lower Amari error than fastICA alone (see Figure 11). In the square case experiments, sparse regularization performed best in the gamma distribution case against fastICA alone. In the overcomplete case, regularization again improved source recovery in a majority of sample size tests (see Figure 12). In the gamma, Laplace, and Rayleigh distribution experiments, $\sigma = 0.05$ sparse regularization improved recovered source quality. For a majority of samples sizes in these experiments, $\sigma = 0.05$ outperformed FastICA alone.

3.8 Discussion

An important observation of applying $\hat{\ell_0}$ to source separation is noting where the function attributed weight to values. Figure 13 shows the plot of the derivative function for $\hat{\ell_0}$, where weight is placed on values near zero. The function places equal absolute weight on values equidistant from zero.

The intuition for regularization using this function is that it places weight on the values that affect sparsity, namely the values near zero. The general form of the function is that it places Gaussian distributed weight on these values. The unbiased nature of the function suggests that using prior information about the data noise model would beneficially impact performance in overcomplete cases where extra sources are not Gaussian distributed.

We have presented an approximation function for regularization of ℓ_0 norm. This function admits unconstrained optimization methods, allowing it to be used in the infomax method of source separation. Experiments with a range of continuous distributions and sample sizes show improved separation performance in cases where noise sources are present and overcomplete instances.

We also show relative performance gains for regularization in RADICAL and fastICA, both popular source separation methods. The comparisons indicate that regularization improves ICA performance in square and overcomplete cases. In the case of infomax ICA, sparse regularization using the $\hat{\ell_0}$ approximate function improves source recovery. In the case of RADICAL, while regularization generally performs as well as or better than RADICAL

alone, the performance gains are not as great as the gains using infomax or fastICA. One reason for this is the discrete search over rotation angles in the RADICAL method. A more resolute search over rotation angles might discover more optimal rotation angles for separation, allowing regularization methods greater freedom to find optimal rotation angles for the combined cost function. The relative performance gains are less clear for fastICA; although sparse regularization does improve source recovery in most test cases.

Finally, regularization improved source recovery in a majority of cases tested, but in $\hat{\ell_0}$, different values of sigma yielded different performance characteristics. Cross validation strategies can be applied to learn optimal parameter settings for sigma. Follow on work pursues learning optimal estimators for sigma values of the approximator given conditions of the data. Gathering summary features of the data before source separation might assist in learning sigma.

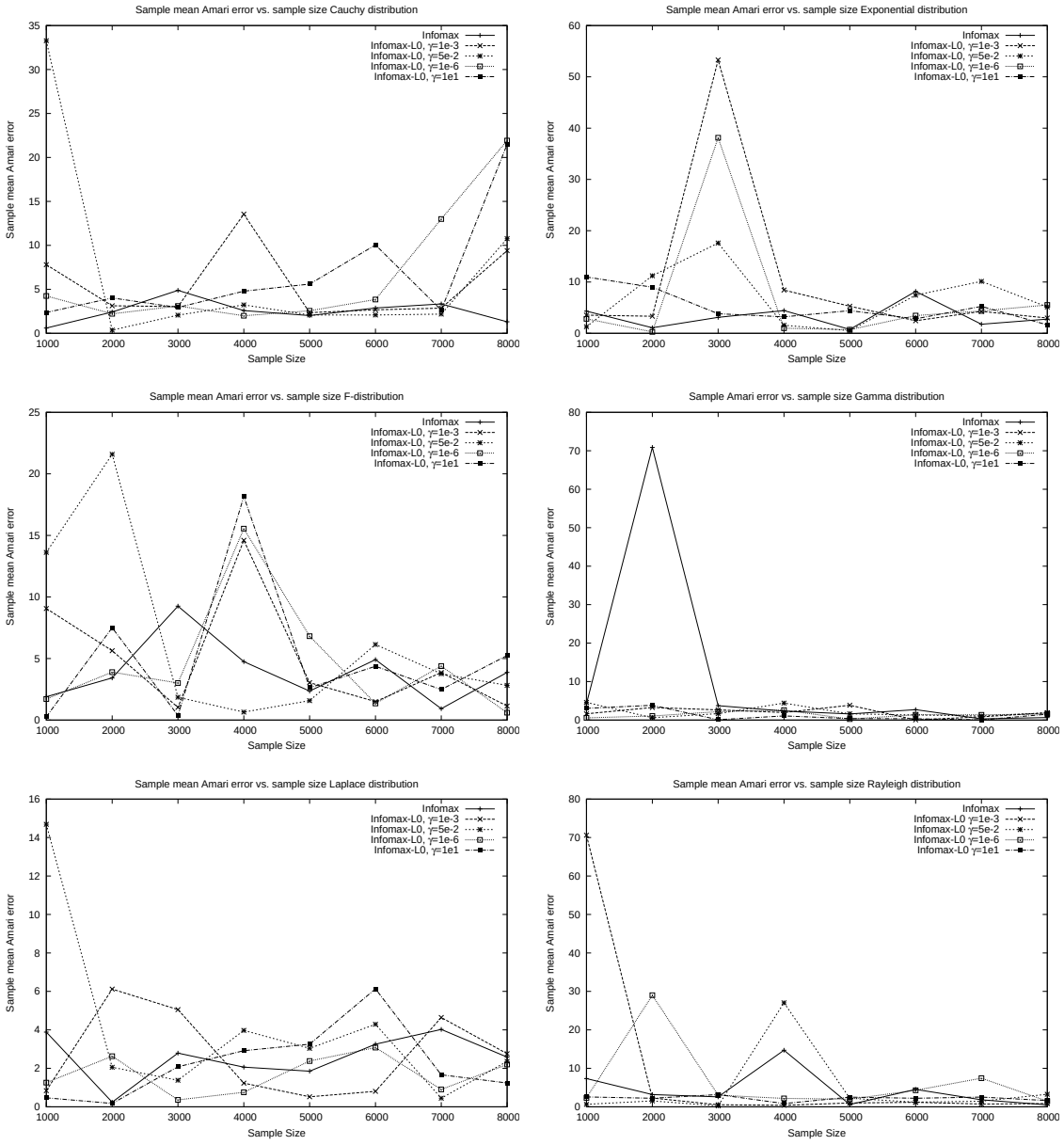

Figure 5: Infomax ICA: Noisy source in square separation results using infomax ICA. Cauchy (top left), Exponential (top center), F-distribution (top right), Gamma (bottom left), Laplace (bottom center), and Rayleigh (bottom right).

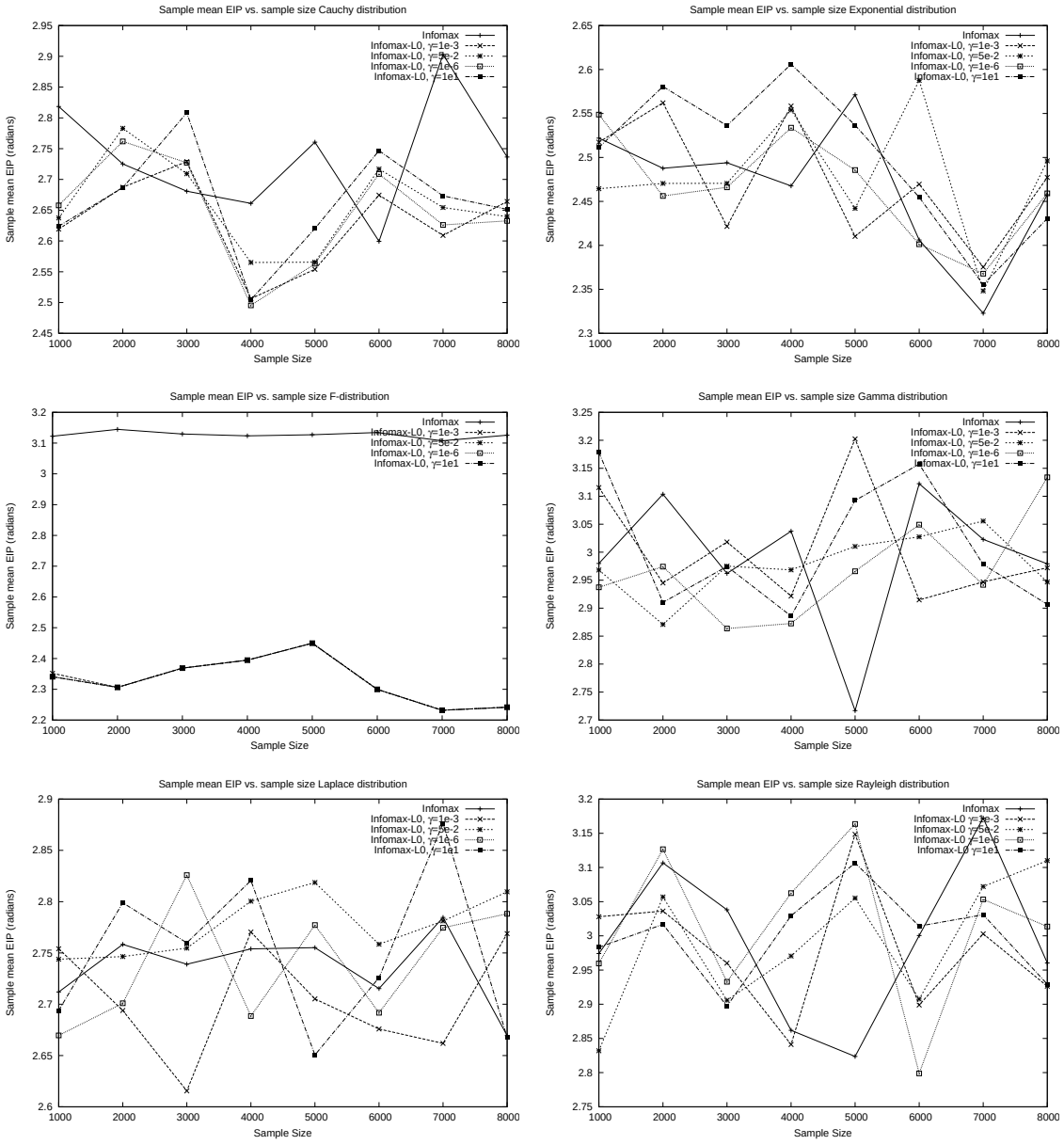

Figure 6: Infomax ICA: sum min EIP (radians) versus sample size for each distribution mixture using infomax ICA. Cauchy (top left), Exponential (top center), F-distribution (top right), Gamma (bottom left), Laplace (bottom center), and Rayleigh (bottom right).

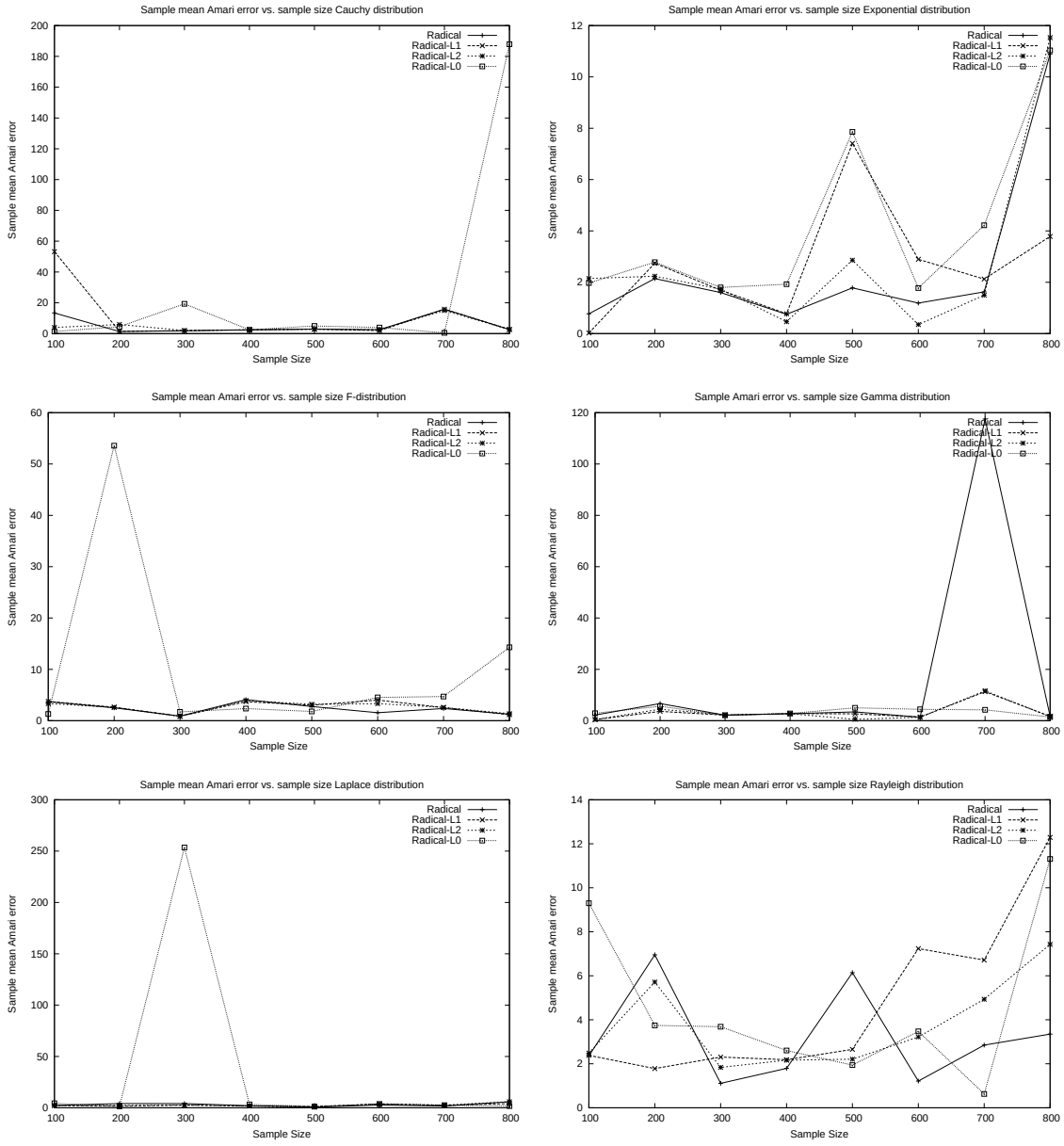

Figure 7: RADICAL Experiment 1: Sample mean Amari error versus sample size for each distribution mixture. Cauchy (top left), Exponential (top center), F-distribution (top right), Gamma (bottom left), Laplace (bottom center), and Rayleigh (bottom right).

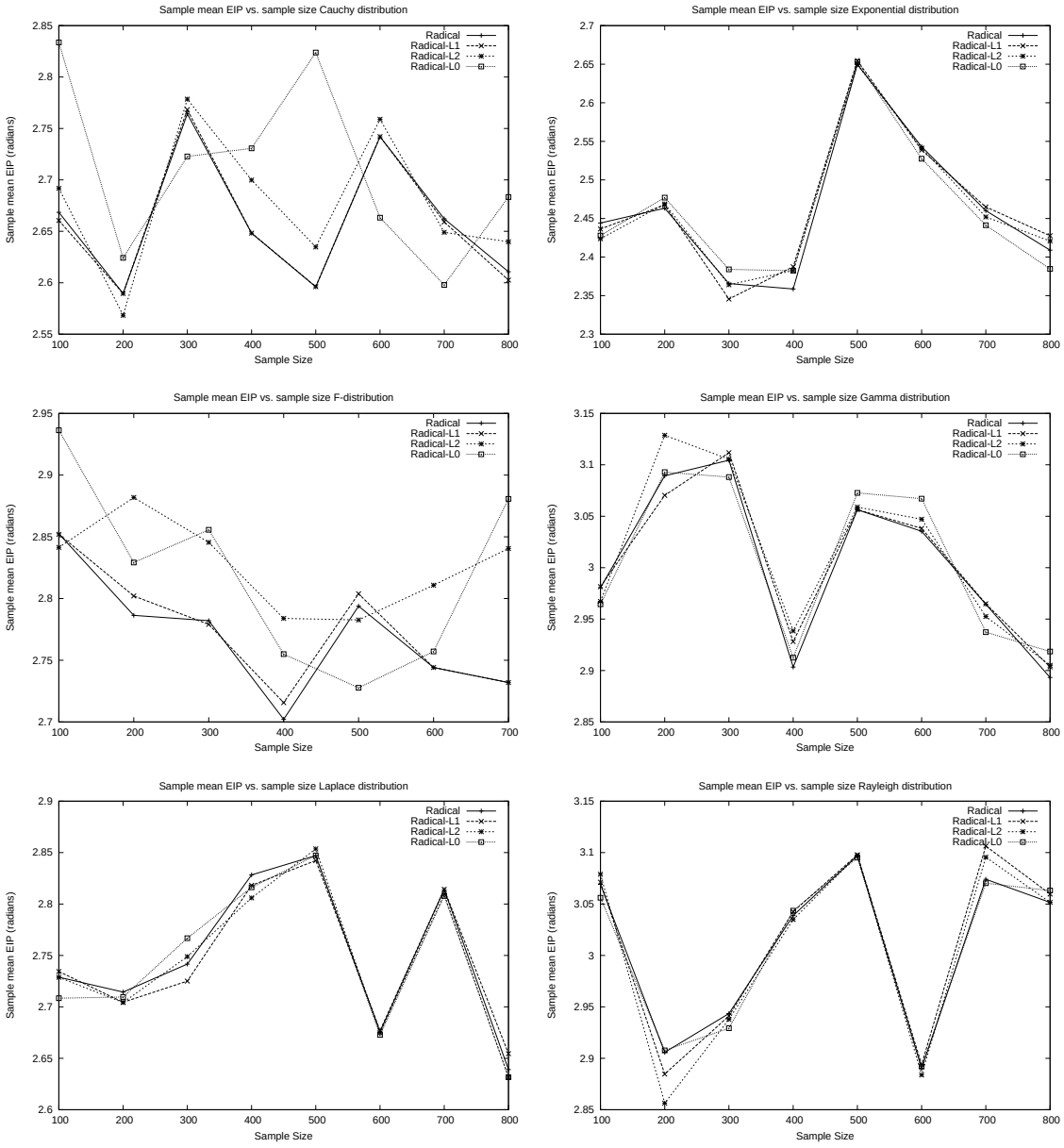

Figure 8: RADICAL Experiment 1: Sample mean min EIP (radians) versus sample size for each distribution mixture. Cauchy (top left), Exponential (top center), F-distribution (top right), Gamma (bottom left), Laplace (bottom center), and Rayleigh (bottom right).

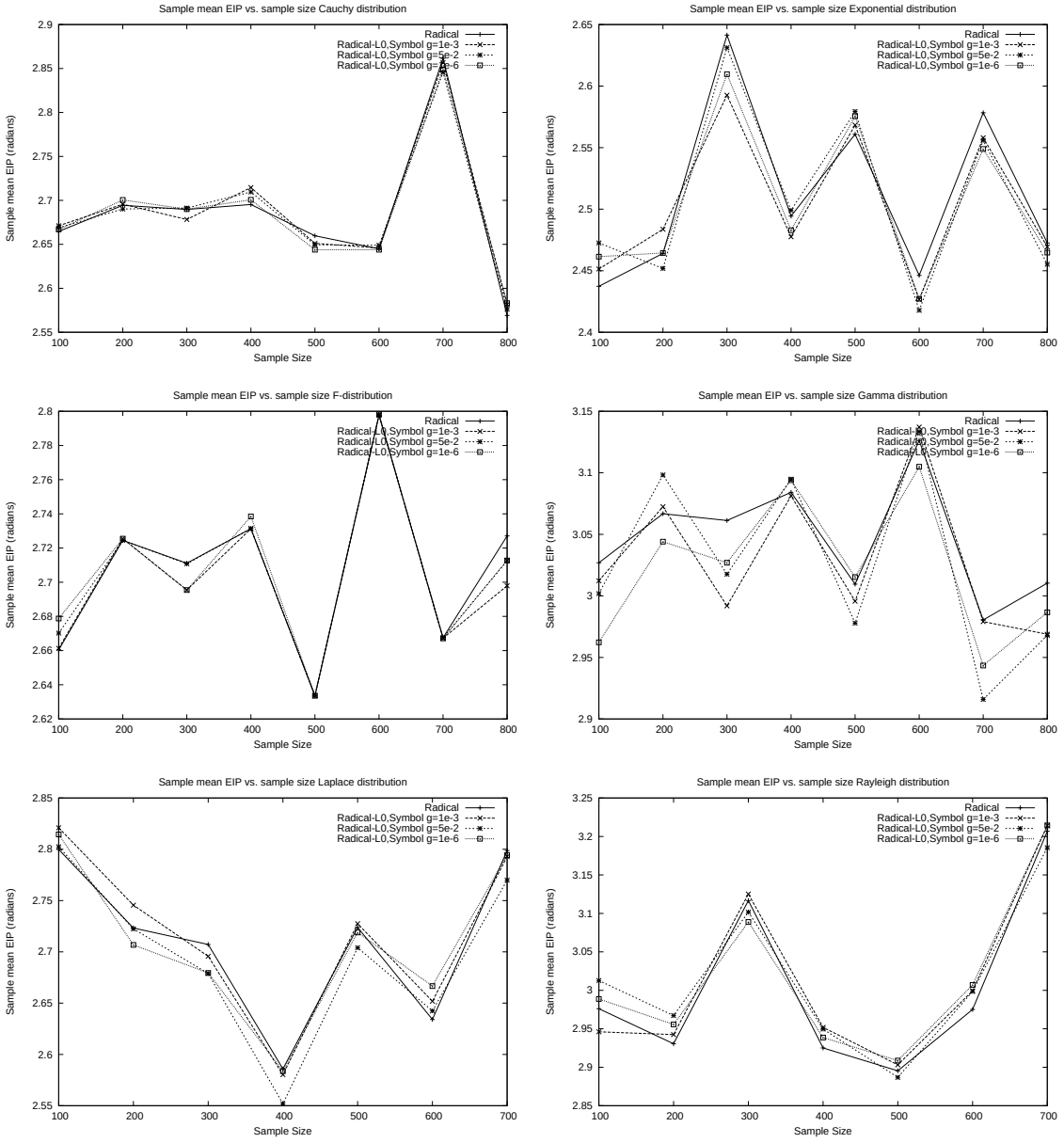

Figure 9: RADICAL Experiment 2: Sample mean min EIP (radians) versus sample size for each distribution mixture using RADICAL. Cauchy (top left), Exponential (top center), F-distribution (top right), Gamma (bottom left), Laplace (bottom center), and Rayleigh (bottom right).

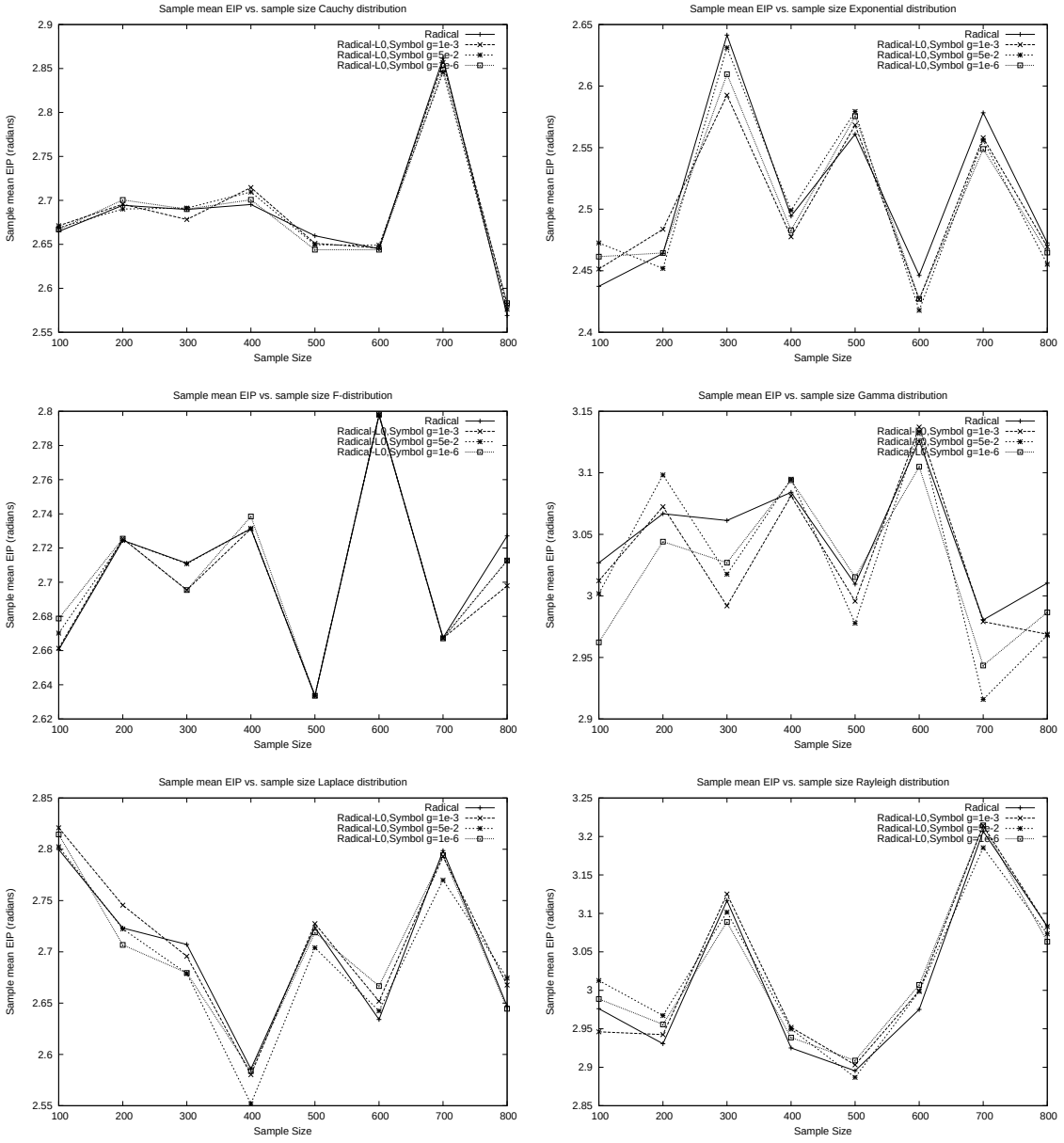

Figure 10: RADICAL Experiment 2: Sample mean min EIP (radians) versus sample size for each distribution mixture using RADICAL. Cauchy (top left), Exponential (top center), F-distribution (top right), Gamma (bottom left), Laplace (bottom center), and Rayleigh (bottom right).

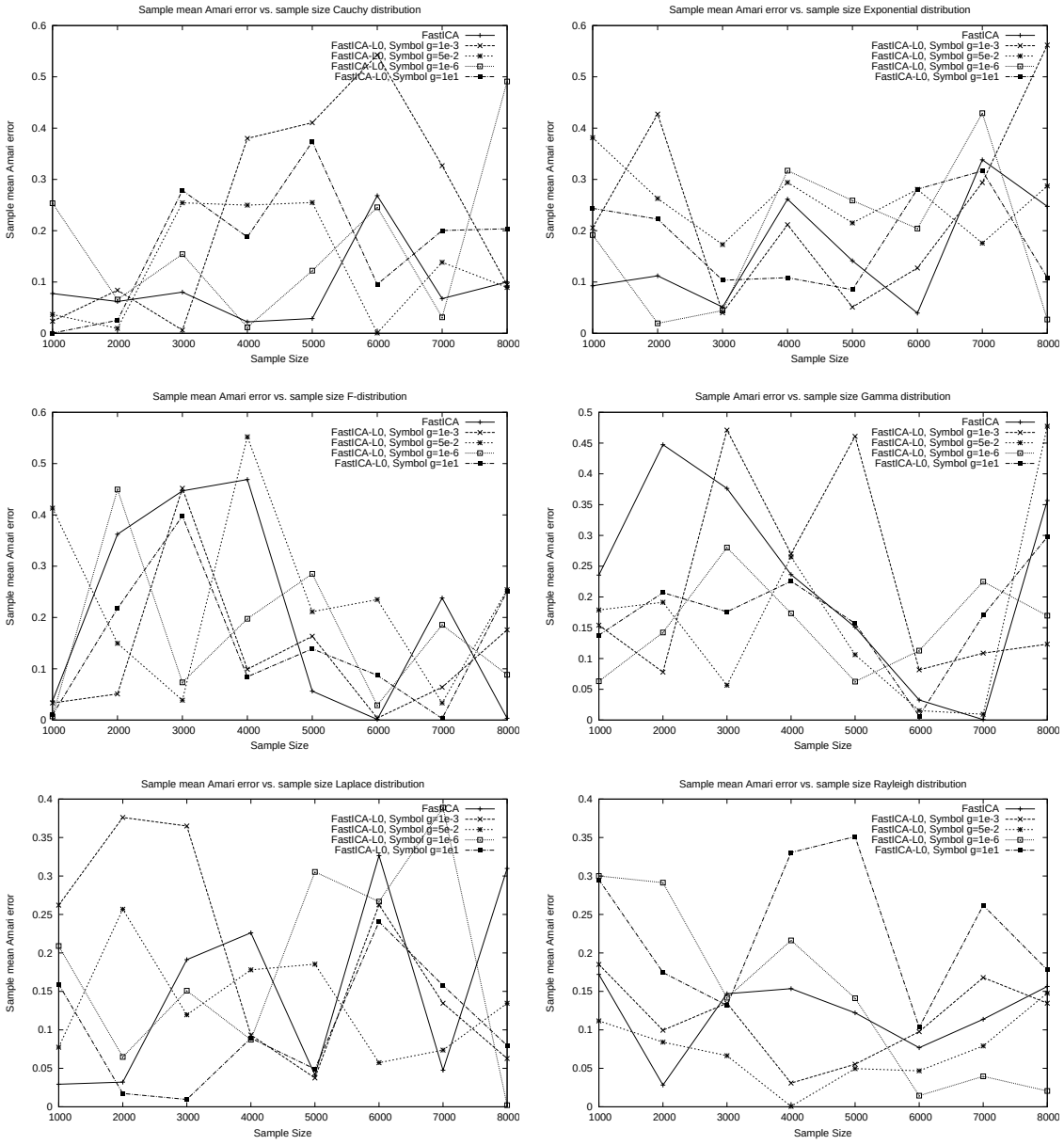

Figure 11: fastICA experiment: Sample mean Amari error versus sample size for each distribution mixture using fastICA. Cauchy (top left), Exponential (top center), F-distribution (top right), Gamma (bottom left), Laplace (bottom center), and Rayleigh (bottom right).

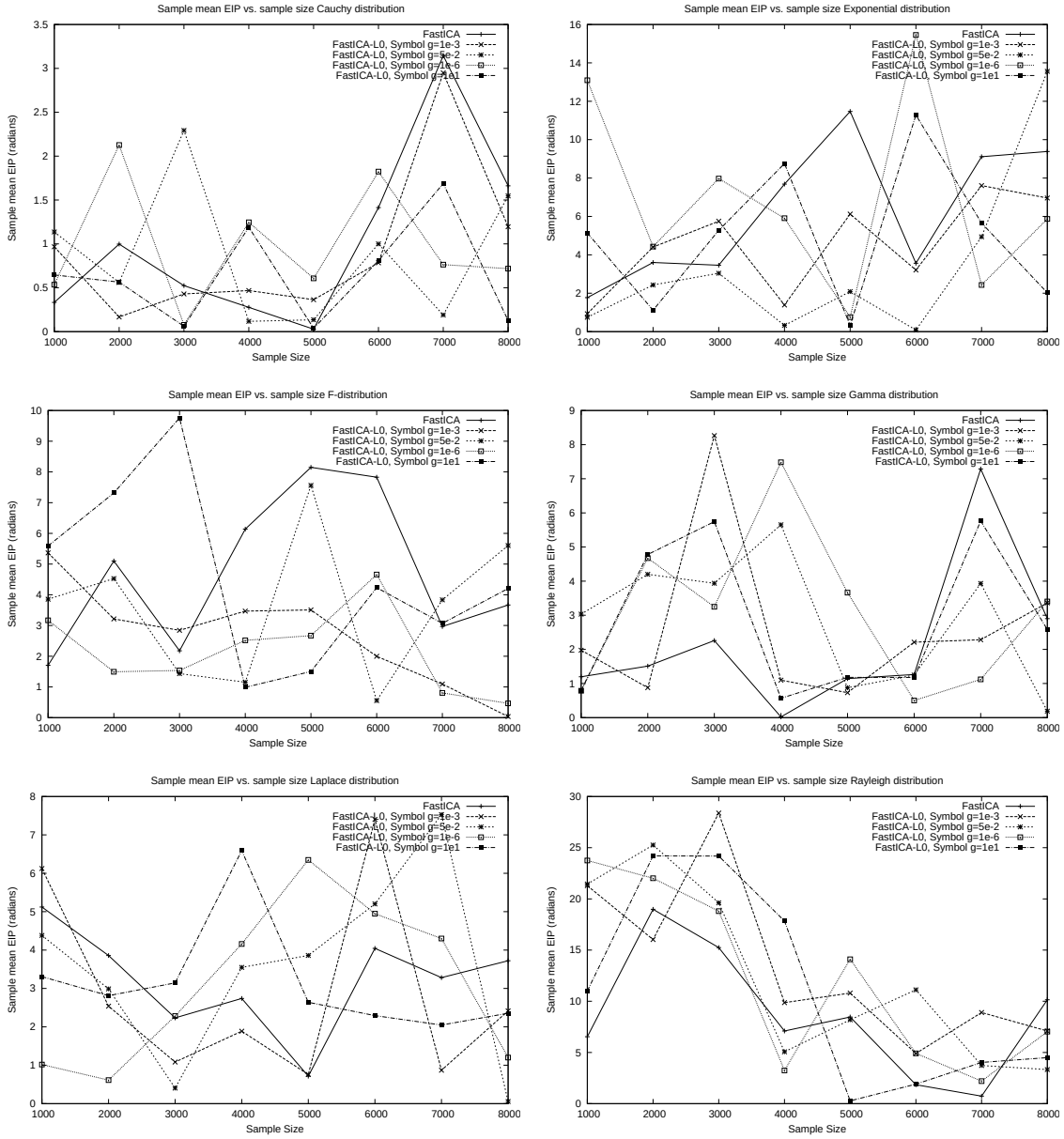

Figure 12: fastICA experiment: Sample mean min EIP (radians) versus sample size for each distribution mixture using fastICA. Cauchy (top left), Exponential (top center), F-distribution (top right), Gamma (bottom left), Laplace (bottom center), and Rayleigh (bottom right).

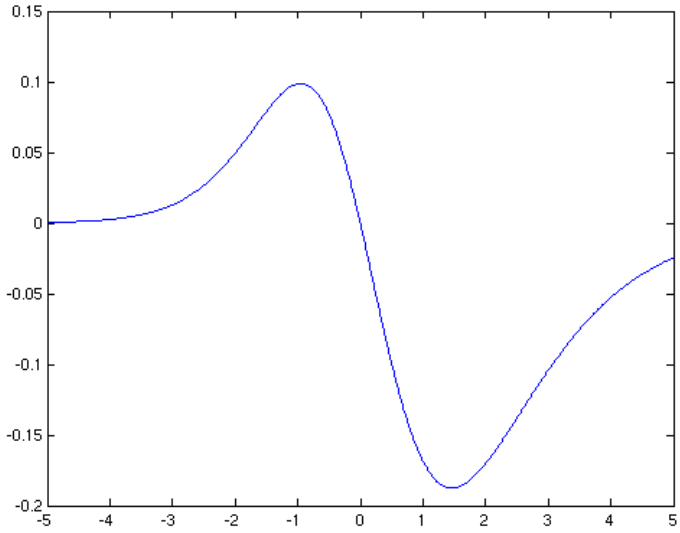

Figure 13: $\hat{\ell_0}$ properties: $\frac{df}{du}$. Regularization weight is placed on values near zero.

CHAPTER IV

MENTAL ROTATION ERP FOR BMI

Mental rotation is a well-studied cognitive phenomenon that has been applied to BMIs (28) as a contrasting task to other cognitive tasks for control. Today, functional imaging methods are revealing more detail of task-related activation for mental rotation. Functional imaging findings suggest that ERP of mental rotation tasks are differentiable based on the task rotation angle. The objective of this work is to study the differentiation of ERPs of mental rotation tasks according to rotation angle of stimuli and suggest that mental rotation tasks are good candidate tasks for BMIs. We approach the study of mental rotation from a Psychophysiological perspective.

4.1 Psychophysiology of mental rotation: implications for brain-machine interfaces

For over thirty years, mental rotation has been the central focus of a number of fundamental debates in Psychophysiology: the representation of imagery of motion contains elements of a number questions important to Psychophysical research. Researchers interested in vision perception models of motion processing are familiar with results on mental rotation. Research in spatial information processing focuses on in functional models of rotation tasks. Researchers interested in motor planning tasks are familiar with mental rotation. Object recognition researchers, one of the fundamental questions in psychophysical research, are interested in mental rotation questions: rotate to recognize and viewpoint oriented vision models all focus on mental rotation.

While many questions of structure from motion representation remain unanswered, there can be little debate that parietal networks play a role in processing motion information, both perceptual and imagery. In particular, networks superior and inferior to the intraparietal sulcus (IPS) region seems to play a major role in these functional tasks. Today, detailed questions of the role parietal networks play in spatial information processing are being

investigated, and a number of debates are being addressed:

- **What is the nature of lateralized effects in rotation tasks?** Previous results have suggested a right lateral bias of rotation tasks, while others show less of a lateralized effect. What are the differences that suggest an explanation for lateralized effects?

- **Are there population-specific differences in the organization and use of intraparietal networks?** Are there differences of the functional organization of mental rotation networks for subject populations?

- **What role do motor processes have in mental rotation?** What is the model of connection of between posterior frontal and parietal? What is the nature of the influence of motor cortex activation in rotation tasks?

- **Where does the dorsal pathway end?** What is the role of parietal networks in motion processing? Mental rotation involves to some extent manipulating objects and conjecture on their path to a destination point in space. Specifically, what is the functional and anatomical boundary of mental rotation tasks?

Current work aimed at resolving these debates is focused on discovering latent structure around IPS for rotation tasks. These experiments are aimed at defining the functional parcellation of the networks around the IPS.

4.1.1 Primate IPS

An important line of investigation in the discovery of latent structure within the human IPS has been electrophysiology studies in primate IPS. The human IPS separates the inferior and superior regions of the parietal lobe. The anterior portion of the IPS is proximate to temporal motor regions, while the caudal portion is proximate to the occipital lobe. In primates, the IPS has a similar location, and therefore it is believed to contain similar functional aspects of motion representation. One trend of this work has been to first find rotation selective responses in primate networks, and then to find human correlates of

these phenomena. Early work using microelectrode arrays focused on finding orientation selective networks in response to simple motion stimuli(30). Microelectrode recordings over the parietal lobe in primates established the role of PPC networks in representing motion, and Desimone and Ungerleider (30) mapped many orientation selective sites in the caudal IPS.

Work directed on the caudate lateral bank of the IPS (c-IPS) was aimed at dissociating motion processing attributed to MT/MTS in primates and object rotation. These studies were motivated to model how the brain represents three-dimensional objects in manipulation tasks and use oriented graspable objects to evoke responses in the parietal regions(98). In these studies, single cell recording revealed responses to a number of graspable stimuli, rotated in the x,y, and z planes. In these cases parietal responses increased under stereo vision conditions versus single eye conditions. Further, graspable objects showed oriented based responses. Harris et al. (50) notes an earlier Sakata study that points out the orientation-selective neurons for hand manipulation tasks in the anterior IPS. Sakata et al. (98) also found in primate c-IPS axis-orientation-selective neurons as well as surface-orientation-selective neurons. For three dimensional objects, neurons responded differently to orientation differences.

Recent studies are aimed at finding human correlates to the relatively well defined structures in primates. Grefkes and Fink (48) presents a summary of data drawn from studies of electrophysiology in the intraparietal sulcus of primates. From these studies, five functional areas are delineated along with their functional putative human correlates: anterior, ventral, medial, lateral, and caudal regions. Anterior IPS responds size and shape orientation of objects. Ventral IPS responds to polymodal input: visual, tactile, vestibular, and auditory stimuli. The putative coarse analog is the PPC in humans. Medial IPS responds to motor planning, execution, and maintenance of reaching tasks. This is also found in the medial IPS in humans, and is ofter referred to as the parietal reaching region. The lateral IPS responds to saccades and eye movements; in humans there is evidence that the analog region responds to attention in the context of controlling eye movements. Finally, the caudal IPS responds to surface and pattern orientation in both humans and primates.

4.1.2 Behavioral aspects

Early work in rotation began with investigations of parietal legion patients who retained normal vision, but displayed difficulty manipulating spatial objects. Subsequent investigation showed these subjects had varying difficulty mentally manipulating objects. The problem with this line of investigation was that researchers were not able to ascertain the functional or anatomical boundaries of these legions. The first behavioral work began with Shepard and Metzler (100). Three dimensional wireframe shapes were used in a rotation task. These wireframe shapes were random configurations of wireframe cubes attached on their faces. These shapes were shown in pairs, and the task was to judge whether the shapes were the same rotated or if they were mirror images of each other. In all the paired cases, the shapes were rotated in the x,y,or z plane. Subject response time increased with the amount of rotation. Follow on behavioral studies confirmed this result using different types of stimuli, most notably, alphanumeric characters(21). Orientation judgment tasks using clockface stimuli also showed increased RT for increased rotation angle(23). Corballis and McMaster (24) showed similar trends in letter-rotation tasks even if the judgement was a left-right decision. In this task, participants were asked to indicate whether a dot displayed with the letter stimulus would be on the left or right of the upright letter.

The improvement of measuring response time accurately, improvements to stimulus presentation methods, and improved imaging techniques have all contributed to a number of findings about the functional nature of mental rotation. Wraga et al. (118) show response time differences for specific viewer perspective related rotations, suggesting the use of frame-of-reference information in spatial manipulations.

4.1.3 Mental rotation: Psychophysiology

Psychophysical work in mental rotation focuses on the behavioral, ERP, and BOLD response differences to different types of rotation stimuli common in the physical environment. The goal is to recover models of rotation tasks to better understand motion representation in the visual cortex. Early imaging work was motivated to find task-relevant regions of activation for mental rotation. PET scan data of mental rotation tasks revealed a localized area in

the right parietal lobe responsible for various rotation tasks (18), using the same stimuli used by Shepard and Metzler (100). Alivisatos and Petrides (3) used fMRI with rotated alphanumeric stimuli to show parietal responses to rotation tasks. Muthukumaraswamy et al. (81) used EEG to contrast ERPs of mental rotation and size transformations. In this study, random generated shapes were manipulated for rotation and relative size (zoom). Barnes et al. (7) contrasts rotation tasks with linear transformation tasks in fMRI.

Follow on studies have addressed aspects of the design of this first experiment, and extended the results to a number of different stimuli. A one criticism of the experimental design used by Shepard and Metzler (100) is with the presentation of the paired images simultaneously until judgment. The result that response increased with the rotation angle in this task might be explained by an increased number of eye movements required to make the comparison (3). Podzebenko et al. (92) presents a coarse parcellation of human IPS regions using fMRI and letter rotation stimuli. Neuroimaging studies have become the standard experimentation tool for more accurate models of mental rotation (44; 7; 50; 111; 59; 53). Gauthier et al. (44) examine object recognition as a function of viewer perspective and the participation of mental rotation in recognition tasks. Heil and Jansen-Osmann (53) examines lateralization effects in children for rotation tasks.

4.2 Task decomposition

Functional perspectives of mental rotation in evoked responses identify several subprocesses to rotation (69):

- Generation
- Maintenance
- Inspection
- Transformation

This task decomposition offers a set of explanations for the variance in results involving mental rotation. An interesting set of experiments would be to ascribe these functions onto

PPC and visual cortex regions, and to contrast those results with visually impaired subjects: cortically blind and degenerative disease subjects.

4.2.1 Debate 1: Lateral effects

In many of these results, right lateralized effects of rotation were shown (89). Corballis (25) summarized the experimental evidence for and against a lateralized effect for mental rotation and posited that while there seems to be a lateralized effect, it is not on the order of language lateralization and therefore not a significant effect. Subsequent explanations have suggested a learning effect in rotation tasks that causes the lateral response: expert rotation has a stronger right lateral response (Voyer, 1995). Today, additional evidence supports both sides of the debate. Harris et al. (50), Alivisatos and Petrides (3), and Ditunno and Mann (32) all support lateralized effects. Revisiting these experiments with a functional parcellation of IPS in both hemispheres should provide a more satisfying explanation.

4.2.2 Debate 2: Population specific differences

One of the more popular claims of mental rotation is that there are population specific differences in response. Jordan et al. (59) showed evidence of greater ventral pathway activation in the case of women, and greater PPC activation for women in rotation tasks. In light of a functional decomposition and parcellation of the IPS region, these results should be revisited. The results fail to pinpoint in the functional decomposition when the ventral pathway activation occurs, and fails to indicate within IPS specific differences in the BOLD time course.

4.2.3 Debate 3: Temporal, frontal influences on rotation

There have been some results about the influences in certain types of stimuli on specific regions of IPS. Sakata et al. (98) used graspable stimuli to evoke responses in primate anterior IPS. The motivation for these studies is to examine the influence of temporal regions on PPC processing; suggesting a top-down model of influence that works in conjunction with the bottom-up model based on the dorsal pathway.

4.2.4 Debate 4: Dorsal pathway

What is the functional and anatomical end of the dorsal pathway? There is a wealth of knowledge that the temporal and parietal lobes contain networks that are relevant or that are employed by visual cortex the for the purpose of representation. These networks are often orientation selective they are likely closely tied to. Defining refined tests for orientation selective responses in IPS not only provides a more resolute parcellation of PPC, but also informs this debate as well in section 4.2.3.

4.2.5 Conclusions

It seems clear now that the superior and inferior banks of the IPS play a role in rotation tasks in both hemispheres, and that sex, while influencing performance to do with the task decomposition, has little to do with IPS activation responses to rotation tasks. Going forward, imaging and ERP studies should take into account the specializations of the IPS within both hemispheres, and put into context the stimuli used to evoke responses. Specifically, lateral differences are important with respect to the type of stimuli (two-dimensional, three-dimensional), their orientation, and practice effects. Population differences should be considered with respect to this functional parcellation: discovering differences in specific regions of PPC better addresses the debate than current models. Finally, modeling the interaction of top-down influences and the bottom-up responses in PPC will address this important debate. Imaging work with fMRI and mental rotation should focus on finding these relative responses to orientation and surface effects in the context of motion noted in primate visual cortex, and indicated in the wealth of human data currently available.

4.3 Mental rotation motivation

We propose two primary considerations in selecting mental tasks to evoke activation using BMIs: conscious control and speed of inferring intent in the interface. Higher order cognitive tasks are more easily consciously modulated than lower order tasks but are typically slower in practice. Slow detection rates are in part due to the relatively small effect detectable in higher-order tasks, making single-trial classification challenging. Further, these methods

require some amount of training of the subject before they are able to use the system with any accuracy.

SSVEP and P300 responses on the other hand are lower-order evoked responses largely outside of conscious control. BMIs based on these lower-order phenomena are typically faster in throughput rate. Evoked effects from these tasks are noted quickly, usually in the occipital region and lend themselves to frequency space analysis. One drawback to using these methods is that they are responses to qualities of stimuli and therefore are only evoked rather than produced. In the case of SSVEP, distinct oscillating patches are typically placed at discrete target regions of an interface. Ideally, we would like to find mental tasks that offer the speed and single-trial accuracy of SSVEP while still under conscious control. We hypothesize that mental rotation is one such task, and when paired with other tasks that may also be generated rather than evoked, offer a reasonable compromise between conscious control and speed.

Finally, the ERP features used in classification should capture the relative differences between tasks. These experiments are aimed at finding task-related activity for rotation tasks in contrast to language production tasks using the same stimulus. Previous empirical work on mental rotation tasks has alphanumeric stimuli to evoke response (20). We propose the use of letter stimuli to evoke both rotation and language production tasks. The objective is to be able to detect only task-related activity. Using the same stimuli allows us to control for any other factors influencing ERP differences. The remaining differences (aside from noise) should produce an accurate set of features of task-related activity.

4.4 BCI work

Previous psychophysical work has been interested in finding mechanisms responsible for rotation(100); for control we require functional understanding of mental rotation tasks to model effects sufficient for single-trial or near-single-trial learning. We begin with defining task-related differences between word production and mental rotation tasks.

Historically for BMIs, mental rotation has served as one of a set of orthogonal tasks that may be used in control problems. For instance, del R. Millan et al. (28) use mental

rotation as the mental task indicating intent of a left turn. We use a similar strategy, beginning by examining the task-related differences between word production and mental rotation in left inferior frontal and parietal activation. In the case of rotation tasks and parietal activation, one consideration we need to account for is the possible top-down frontal or temporal influences on rotation, and the interaction of these influences with language production tasks(111).

4.5 Mental rotation versus language production

Language production refers to the broad set of tasks involving generation of language (58). We focus on subvocal word generation for its ERP properties as well as its suitability for control. Like mental rotation, subvocal language production has been the focus of Psychophysiological and Cognitive Science research for some time and remains an active area of research (70; 35).

Subvocal language generation is a lateralized response; its ERP is most easily detected in the left inferior frontal cortex, the region typically referred to as Broca's area. In general results of activation correlated with language production are typically found in the latter ERP waveform (70). Many studies focusing on production and oddball language tasks find task-related activation in the N400 and P600 ERP components (40; 42). Functional imaging studies also indicate that different production tasks yield different activation patterns in Broca's area (41).

We contrast rotation task-related activation with covert word generation task-related activation. We evoke language production using the same stimuli to evoke rotation activation: the rotated letter can either be rotated or used as the start letter in a word generation task.

4.6 Experiments

The converging evidence of functional parcellation of the human posterior parietal lobe suggests evoked rotation perception for objects rotated in the x and z planes will produce differentiable ERPs. More specifically, activation differences at the intraparietal sulcus evoked from manipulations of rotation axis and angle should be detectable using ERP.

The experiments described here are meant to test this hypothesis. The prediction is that different regions of IPS will respond to different rotation axes as well as relative rotation angles within each axis. This is manifest in differences in the N150 and N2 responses.

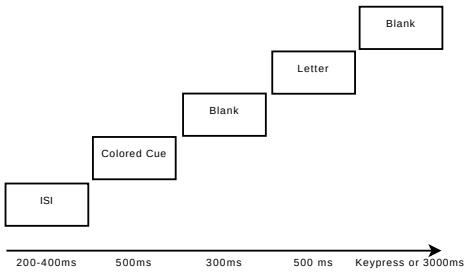

Figure 14: General trial order for experiments using mental rotation tasks.

4.6.1 ERP analysis methods

Previous early work using contrasted two-dimensional and three-dimensional rotation tasks used a power spectral analysis method to generate frequency space features for classifying user performance of the rotation tasks(74). More recent work using mental rotation in general as a contrast task to other cognitive activities reported better results of recognizing rotation in general, but not specific types of rotation. In this work, we analyze evoked response potentials by averaging ERP waveforms of rotation trials and compare to averaged waveforms of word production trials.

4.6.2 General Methodology

In each of the experiments, the same general paradigm was used to present stimuli. A 21in. CRT monitor was used to present stimuli, at 1600x1200 resolution. For all stimuli, a black background was used to display stimuli. Participants used a standard PC keyboard to indicate responses when prompted.

Stimuli were presented using a uniform presentation. All stimuli followed the same presentation sequence. A colored box, 150x150 pixels was shown centrally for 500ms; the colors were red, green, and blue (0xFF0000, 0x00FF00, 0x0000FF respectively). An alphabetic character was then shown centrally for 500ms. The activation caused by viewing the colored cue is distinct from the activation caused by the letter stimulus, and does affect the

activation measured after the offset of the letter stimulus (109). The character was in a particular font and a particular point size. Images of the characters were rendered using the Gimp graphics software package. The images were rotated in the x and z planes in 30 degree increments. A random sequence of rotated character images was generated and shown to participants. In the two studies, participants were instructed to perform specific imagery tasks in each trial, depending on the color of the initial box. We used EEGLAB as the primary tools for ERP analysis (29).

4.7 Rotation pilot studies

We collected data from three subjects in pilot study training. This data served to provide initial evidence validating the experimental design(76). The data also showed no significant differences in the early components of the averaged ERP for the parietal channels. Pilot data also suggested that centered stimuli were needed to remove eye movement artifacts present in the signals as a result of the ordinal located stimuli. In terms of usability, an interface that uses spatial regions of different rotation angles has to account for eye movements in the early components of the ERP.

The latter components for rotation tasks are better targets for amplitude difference analysis, particularly with respect to spatiotemporal analysis using the scalp distribution projections. Heil (52) shows latter component differences in rotation tasks correlated with rotation angle (see Figure 15).

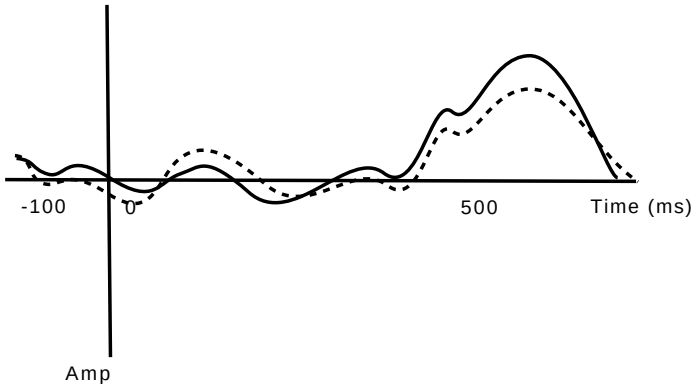

Figure 15: Idealized P500 response for rotation tasks. A decreased positivity response is correlated with increasing angle from 0-180 degrees.

4.7.1 Experimental setup

We recorded behavioral responses and ERP using EEG where the user task is to estimate if the upright figure was valid or not. Invalid characters consisted of backwards or mirror-symmetric characters (i.e. rotated 180 degrees about their central axis). Planar rotations were at 30-degree increments from 0 to 330 degrees. Participants used a keypress to indicate the behavioral response. Characters were placed at the four ordinal locations of the screen. Subjects were shown a cue prior to the stimulus window for 300ms, indicating which of the four stimuli to assess. The ordinal stimulus was shown for 600ms.

4.7.2 Methodology

3 subjects (1 female, mean age 23, all right handed) each completed 288 trials consisting of the ordinal stimuli. ERPs were recorded at 500Hz using 26 sensors arranged in the 10-20 system: Ft7, Fc3, Fcz, Fc4, Ft8, F7, F3, Fz, F4, F8, C3, Cz, C4, Cp3, Cpz, Cp4, C3, Cz, C4, P3, Pz, P4, Oz (see Figure 16). Two reference electrodes on the left and right mastoids and four EOG electrodes were used. Participants were shown four practice trials prior to the experimental trials. Trial data were filtered for artifacts such as blink correction, band pass filtered, and trials were rejected for threshold activation (76).

4.7.3 Discussion

The ERP patterns as a function of rotation angle suggest that the orientation selective networks noted in primate electrophysiology and fMRI based studies of human parietal networks are influencing ERP responses. Being able to detect these patterns in ERP enables BMIs based on rotation tasks alone. These advantages are the motivation for a closer examination of mental rotation tasks.

4.8 Main study

The preliminary results in pilot data suggest mental rotation ERP differences are detectable using relatively few trials. The main study aims are to confirm this claim. The aims of the study are again to first contrast mental rotation task activation with language production

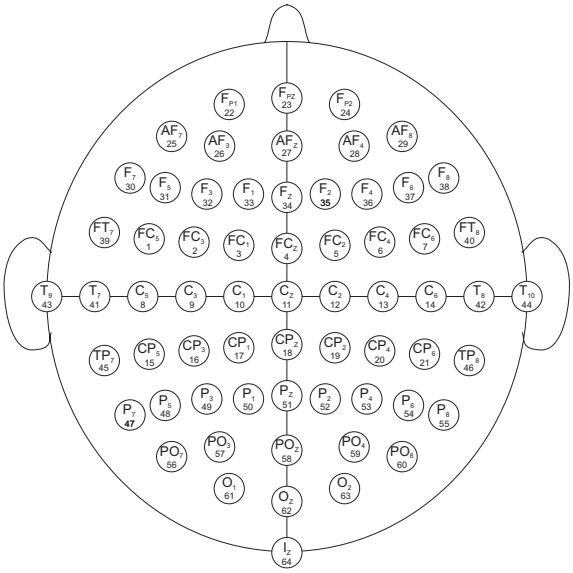

Figure 16: 64 channel montage of the 10-20 placement system (99; 45).

activation. This study also aims to present evidence of recovering orientation selective network activation differences in ERP data.

4.8.1 Methodology

The main study generally uses the same stimulus design as in the pilot studies, where the ordinal stimuli were replaced with centrally located stimuli. The cue colored box is also central onscreen, so the central stimuli removes the necessary eye-movements in the ordinal case.

4.8.2 Study 1: rotation versus word production

We collected ERP data from 15 participants who received course credit for participation. Participants were instructed to rotate the letter stimuli in the first cue (red box) case and to produce nouns and verbs (blue and green respectively) in the other cues. Rotation

judgments were indicated with a keypress and recorded.

4.8.3 Study 2

We collected ERP data from 9 participants who received course credit for participation. In this case, participants were instructed to rotate the letter to its upright position, and evaluate whether the letter was forward in orientation (*i.e.* its canonical depiction) or reversed. Participants responded using a keypress in the red cue condition, and covertly in the blue and green conditions.

4.9 Results

For both datasets, we performed an averaging analysis of the early and late ERP waveforms. We collect trials for all subjects by conditions, and average the time series data to form grand average waveforms of the ERP. For the experiment conditions, noise elements in the prestimulus average introduced significant amplitude components in the averaged waveforms. These components prevent an averaging analysis of single channel data in the array.

4.9.1 Rotation versus word production

Early ERP components. The averaged waveforms of selected left and right parietal channels for word production and rotation tasks. The early components of the data contain noise components, but the main perceptual components of the ERP are present and are loosely coherent between conditions. The early averaged waveforms do not seem to indicate a clear task-related difference in the perceptual components. The intuition for this is that the stimuli were the same for both task conditions, and therefore should not produce significant differences in the early components.

Latter ERP components. The latter components are less correlated than the early, perceptual components.

4.9.2 Covert rotation versus response

The grand average for rotation tasks shows a slight left lateralized effect, where the perceptual components are more stable in left parietal channels than the right channels. Again,

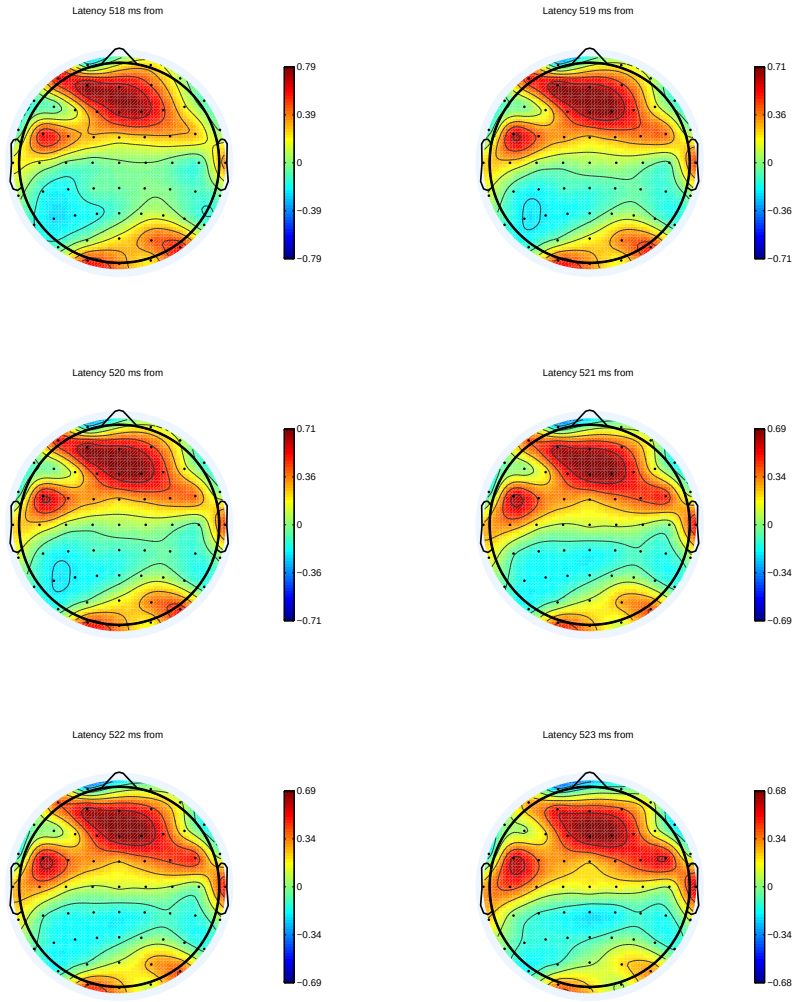

Figure 17: Scalp distribution time course from 518 to 523ms of language task.

noise components present in the prestimulus components of the ERP prevent single channel averaging analysis. In both the language production and the covert response experiments, the scalp distributions indicate activation differences based on region, but in both cases the traditional analysis is not able to corroborate the differences.

4.10 Discussion

We have shown task-related differences in the late components of the ERPs for word production and rotation tasks in the left frontal and left parietal channel data. The averaging analysis indicates a negativity in parietal channels for rotation that corresponds with a

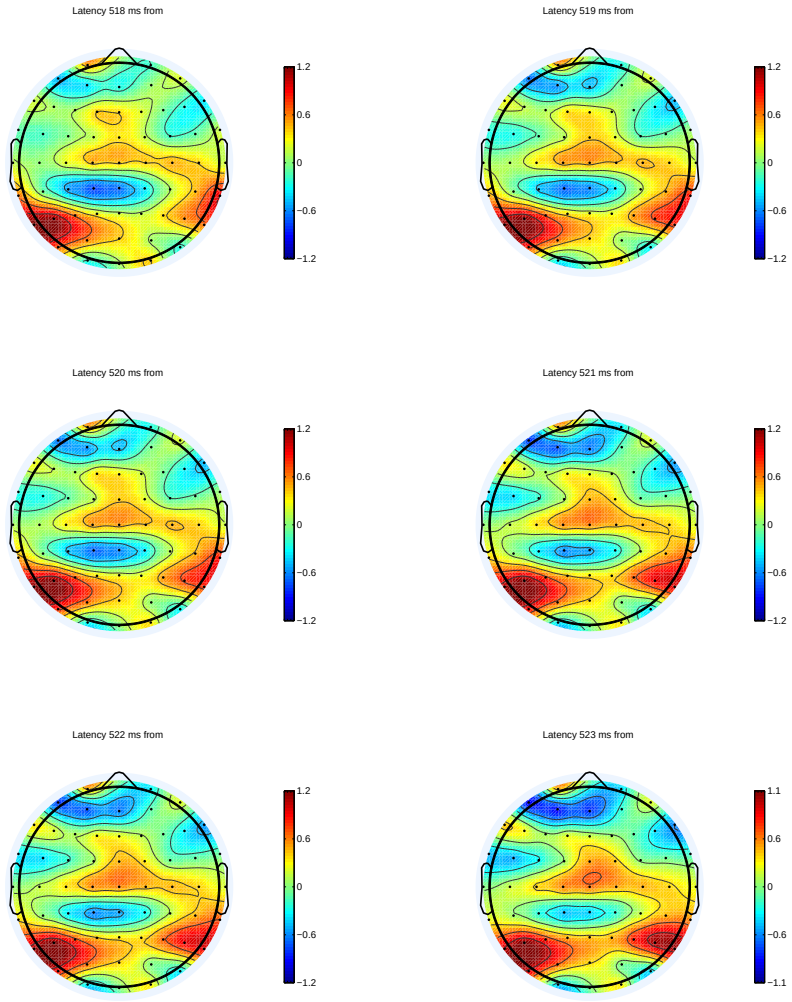

Figure 18: Scalp distribution time course from 518 to 523ms of averaged difference ERPs between rotation and language production tasks.

positivity in left frontal channels for word production at 500ms. Likewise, a positivity in parietal channels corresponds with a negativity in frontal channels for rotation tasks (see Figures 17 and 18).

CHAPTER V

EARLY ERP DIFFERENCES ASSESSING APPROXIMATE SYMMETRY

The role of structural information in early visual processing is a focus of research in functional models of visual perception. The regularity of bilateral symmetry and the sensitvity of the human perceptual system to symmetry through oblique and occluded angles suggests the use of symmetry information in early perceptual objectives like figure ground discernment and grouping. If symmetry detection employs simple operations of symbol representations of visual features, manipulation of structural information in symmetry representations should result in detectable differences of symmetry assessment. In this study, we present empirical evidence of differences in assessment of symmetry based on manipulations of structural information of bilateral symmetry.

5.1 Introduction

The human perceptual system is sensitive to bilateral symmetry in the visual environment (114). Indeed, in objects where only approximate symmetry exists, symmetry assessment is successful even in oblique or occluded angles (107). Early behavioral studies on symmetry noted several interesting effects including fixation bias to one side of symmetric stimuli (73) and lower reaction time (RT) to vertical symmetry assessment over oblique symmetry axes (86) that motivated interest in symmetry assessment research. Two stage models of symmetry processing have been posed to explain behavioral differences in two phenomena: fast symmetry assessment and iterative feature inspection methods (22).

In the early visual perceptual system, where little inferred information about the visual environment is available, fast symmetry assessment seems to be important, supporting grouping and figure-ground objectives. At the same time, information about how features compare requires top-down processing and is too costly in time for its use in representation during early visual processing. From a functional perspective, symmetry assessment requires

feature comparison along a hypothesized axis of symmetry. This study addresses the nature of the use of structural information in early visual perception.

5.1.1 Structural Information Theory and Symmetry

Research in modeling visual shape perception is concerned with discovering how information is encoded and processed in the visaul system. This work takes a functional view of information and treats environment features given by the retina as symbols. Models of visual perception are defined through functional operations of these symbols. Given the timing constraints of early perception, functional operations are thought to be simple, involving few encoding rules, and fast, allowing for parallel operation over the retinal encoding (107).

Structural Information Theory (SIT) formalizes a method for modeling visual perceptual phenomena. SIT treats feature information obtained from the retinal encoding as symbols and defines operations in terms of symbol manipulation (108). Extensions and applications of SIT to modeling grouping meachanisms in early perception used a symbolic symmetry operation where features could be compared about a pivot point (107). Other models of early perception also use structural information as a motivator for their modeling strategy (87).

Encoded visual information is treated as symbols in many perceptual models, and encoding rules are symbol manipulations that result in representations of the visual environment that explain empirical results. The perceptual relevance of symmetry in intermediate representations of objects is prevalent in perception models of early visual processing. The role of symmetry as a fundamental coding operation is at the heart of perceptual models, in part due ta the symbolic representation of features. Symbolic representation of features lends itself to symmetry: a fundamental operation that gives rise to complex structure analysis that has the benefit of parallel processing.

If symbolic representations using symmetry as a primitive operation detect approximate symmetry, then manipulations of the structural information should affect approximate symmetry judgment performance. We present evidence of the perceptual relevance of bilateral

symmetry in early visual processing and explore aspects of bilateral symmetry processing that are contained in early processing. Symmetry differs from other early processing properties in that it involves structural comparisons of opposed features.

5.1.2 Structural asymmetry

Structural information theory makes some predictions about the nature of assessing symmetry in terms of symbol manipulations of the feature representations of figures. Consider a symbol representation of a segment of a polygon contour in assessing symmetry (107). The symbol string used to represent a polygon contains symbols of vertex angles, distances of vertex angles from center, and line segment lengths between vertices. A string of these symbols fully describes a polygon stimulus. To assess symmetry functionally given this representation, a symmetry operator is applied to the symbol string, where features are paired about a pivot (the hypothesized axis of symmetry). The symmetry operator compares feature pairs, where all matched pairs indicates symmetry, and a mismatch indicates symmetry. It is also possible to rank relative symmetry under this model using the number of mismatches that take place.

We define two symmetry violations that represent different changes to structural reprensentation. First, quantitative asymmetry manipulates the location of a polygon vertex, but preserves the convexity quality of the corresponding symmetric vertex. In terms of the symbol string representation, the difference affects three symbols. Second, qualitative asymmetry also manipulates the location of a vertex, but inverts the convexity property of the corresponding symmetric vertex. This manipulation changes an additional symbol in the SIT representation. If structural information in the form of a symbol string is used to assess symmetry, performance should be improved in qualitative cases, with the additional symbol manipulation to improve detection rates (see Figure 19).

5.1.3 Previous Work

Symmetry is a property of contours of objects, and previous behavioral studies focused on relating symmetry to the information content of stimulus contour. Early behavioral work with symmetry suggested evaluation bias to one side of vertically symmetric figures (73).

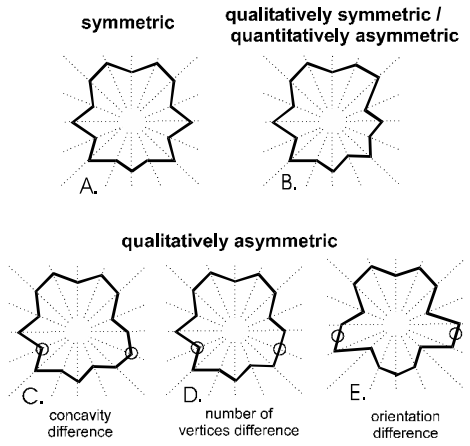

Figure 19: Structural asymmetry differences.

Palmer and Hemenway (86) show that RT increases with increased rotation of the axis of symmetry from vertical. The interpretation of this is that symmetry is most easily detected in the vertical (central) axis of objects. This result compares well with the intuition that most objects in the visual environment possess inexact symmetry in the vertical axis. Locher and Nodine (73) suggest conditions of stimulus symmetry influence behavioral response to judgment tasks when contrasted with repeated contour patterns.

Initial work relating structural information to symmetry evaluation used behavioral measures to demonstrate significant differences in evaluation strategies. Later behavioral studies using eye-tracking suggest significant differences between eye-movement patterns in symmetry evaluation of figures containing structural manipulations; in the face of a general search strategy that starts in the upper left region, scans down the hypothesized axis, and then compares features of interest.

5.1.4 Hypothesis

If structural information is at play in the early evaluation of symmetry, we can test the manipulation of features that are structurally related against performance in symmetry assessment. The early event related potential should tell us the most about the signal. We expect the ERP to yield detectable differences in the time course where grouping mechanisms are normally significant, namely in the P2/N2 regions. If we believe symmetry to

be an early feature of objects used by the visual system, we would also expect temporally early ERP components (before 300ms) to vary according to symmetry conditions.

5.2 Methodology

The experiment should replicate the results of previous work in behavioral responses of symmetry judgments tasks using the same stimulus generation method. In the case of the ERP study, a single size manipulation was used to detect differences.

5.2.1 Capping procedure

We recorded ERP of the posterior region for symmetry judgment tasks using a Neuroscan NuAmps amplifier with 500 Hz sampling rate. The 10/20 format electrode montage used was: P3, Pz, P4, T5, T6, P07, P08, O1, Oz, O2 with EOG and a common average reference on both mastoids. A ground electrode was used anterior to Fz.

5.2.2 Stimuli

We used the method defined in Mappus et al. (75) to generate symmetric and asymmetric stimuli. We generated 200 stimuli, 100 symmetric and 100 asymmetric. Of the asymmetric stimuli, 20 of each of the 5 asymmetry conditions were generated. We rotated stimuli 90 degrees to generate 100 additional stimuli, 180 degrees for 100 more, and 270 degrees for 100 more. We presented 800 stimuli in random order to each participant in two blocks. Stimulus resolution was 800x600. The large and small conditions were generated using discrete area selections using a measure of area given in Zabrodsky et al. (119).

5.2.3 Experiment setup

Experiment software Presentation was used for experimental trials. A fixation cross was shown for a uniformly random amount of time between 200 and 400 ms before stimulus onset. The stimulus was shown 100ms followed by a fixation cross and a "respond" prompt shown until the participant made a keypress selection indicating whether the participant judged the stimulus symmetric ('0') or asymmetric ('1').

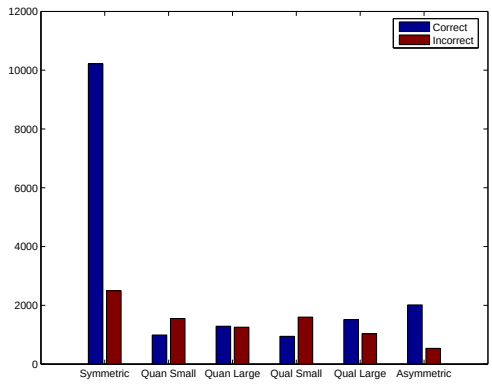

Figure 20: Behavioral data by stimulus condition: Trials counts are plotted according to stimulus condition.

5.3 Results

We decompose the initial analysis by symmetry and size conditions. In waveform comparisons, we generate the grand sample mean of the conditions compared to demonstrate differences. Repeated trials ANOVA are used to determine significance ($p < 0.01$) where differences are noted. Segments of the various ERP components are generated using the super grand mean waveform, and plotting the x-axis crossings for

Scalp distributions showed the P1 and P2 effects, in the occipital channels, but did not indicate particular differences between correct and incorrect responses.

5.3.1 Correct vs. Incorrect

Participants who performed above chance in the symmetric and asymmetric conditions were used in the analysis of results. Participants performed best in these conditions, and demonstrate task attention throughout the process. Behavior response data indicate the relative difficulty of stimulus conditions. The large difference conditions responses indicate high detection rates, while the small conditions are at or below chance. Accuracy was high for both symmetry and total asymmetry conditions (see Figure 20).

5.3.2 Symmetry vs. Asymmetry

We first compare the ERP in two conditions: symmetry and total asymmetry. Participants performed best in these two conditions (highest judgment accuracy); indicating participant

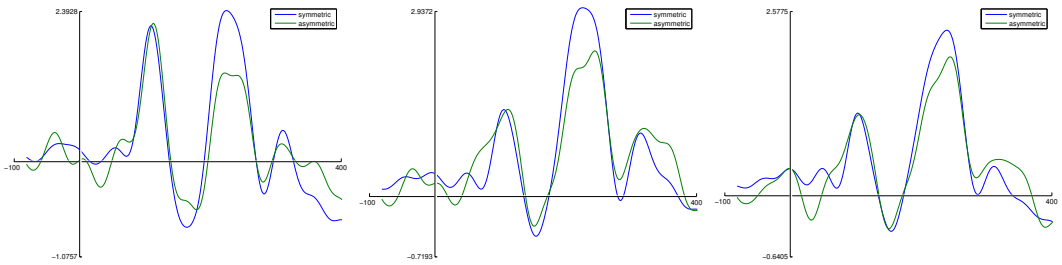

Figure 21: Symmetry ERP vs. Asymmetry ERP T6 (left), P3 (center), and Po1 (right) each show P2 differences for symmetry vs. asymmetry.

attention to the task was sufficient for determining significant differences between ERPs between conditions. From this comparison, we make some general observations that provide intuition for responses in the structural asymmetry conditions. We focus analysis on P1 and P2, because for many channels in the array, these segments represent the most significant differences over all conditions (see Figure 21).

Between symmetry and asymmetry, there were no significant differences in P1. In P2, we note that the grand average symmetry amplitude is greater than asymmetry. The grand average in this experiment represents the average of the recorded time series for all trials from all subjects for a stimulus condition. Further, P2 for incorrect asymmetry responses is less than the symmetry correct responses. The amplitude at P2 may reflect perceived, although that is not indicated in the amplitude in P2 for incorrect responses in symmetry and asymmetry.

One point is that there is a large amplitude difference in P2 for correct responses in symmetric and asymmetric conditions. These are the two conditions to which participants responded with highest accuracy. These differences are some indication of perceived symmetry, and in structural asymmetry conditions serve as the indicator for perceived symmetry.

5.3.3 Quantitative Differences

The quantitative structural asymmetry condition detection time course is best shown in the P3 channel. P1 and P2 amplitude differences are significant in P3 and show differences between correct and incorrent responses (see Figure 22). For correct responses, P1 amplitude is similar in most channels to the symmetric and asymmetric conditions. For

incorrect responses, P1 amplitude is significantly higher than correct response symmetric and asymmetric, and the same is true for incorrect response symmetric and asymmetric.

In P2 amplitude, correct response amplitude matches correct response asymmetry responses. In incorrect responses, P2 amplitude matches correct response symmetry. This trend is noted throughout the structural asymmetry conditions, suggesting that P2 amplitude response indicates participant perceived symmetry condition. While these trends are best noted in different channels for each structural asymmetry condition, the trend is detectable in Po1, P3, and Pz for structural asymmetry conditions.

For large differences, participant response accuracy was higher in the large quantitative condition than small quantitative. The P1 response amplitude is similar to correct response symmetry and asymmetric conditions, while in incorrect trials, P1 amplitude response is higher than symmetry and asymmetry in either correct or incorrect responses. In large differences P2 amplitude is again higher for correct responses than incorrect for selected channels. Again, P2 amplitude differences for correct and incorrect responses reflects the perceived symmetry response.

5.3.4 Qualitative Differences

The quantitative structural asymmetry condition detection time course is best shown in Po1 (see Figure 23). In the qualitative structural asymmetries, P1 amplitude is higher than corresponding quantitative asymmetries for both correct and incorrect responses. For in the small condition, again P2 amplitude response indicates perceived symmetry, closer to asymmetry in correct responses and closer to symmetry in incorrect responses. For large differences, the P1 response is consistent with the qualitative small condition. As with the other structural asymmetry conditions, P2 amplitude responses indicate perceived symmetry conditions.

5.4 Discussion

P1 differences between correct and incorrect responses across the stimulus conditions suggests early ERP captures early symmetry assessment. P2 differences seem to indicate perceived symmetry; in general, lower amplitude in P2 indicates asymmetry response, while

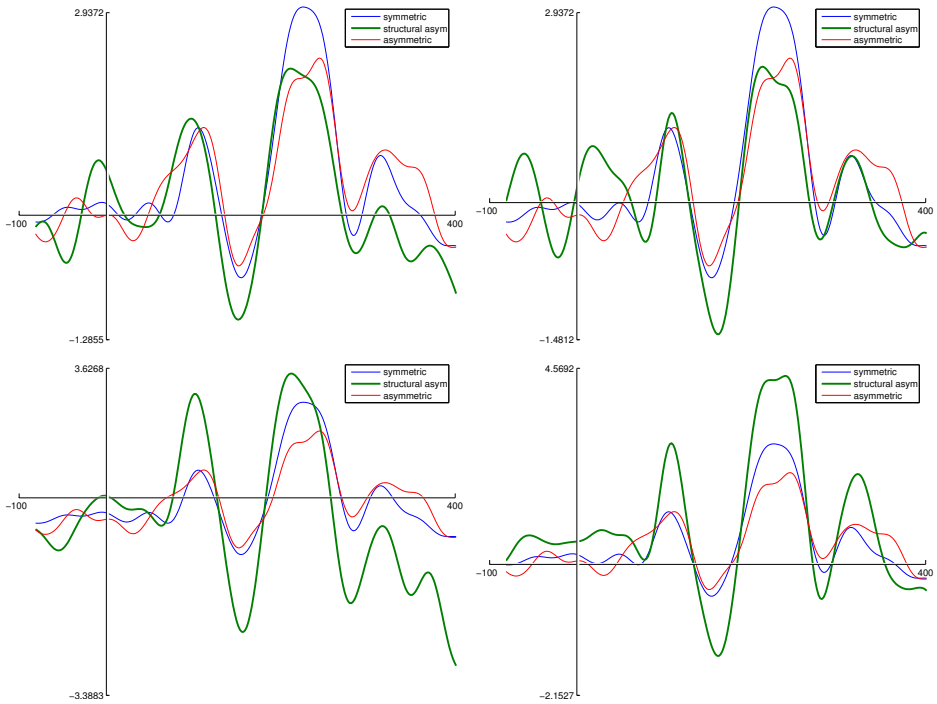

Figure 22: P3 ERP differences of quantitative asymmetries against symmetry and asymmetry conditions: correct (top) and incorrect (bottom) responses. Both small differences (left) and large differences (right) suggest P2 activation indicates perceived symmetry.

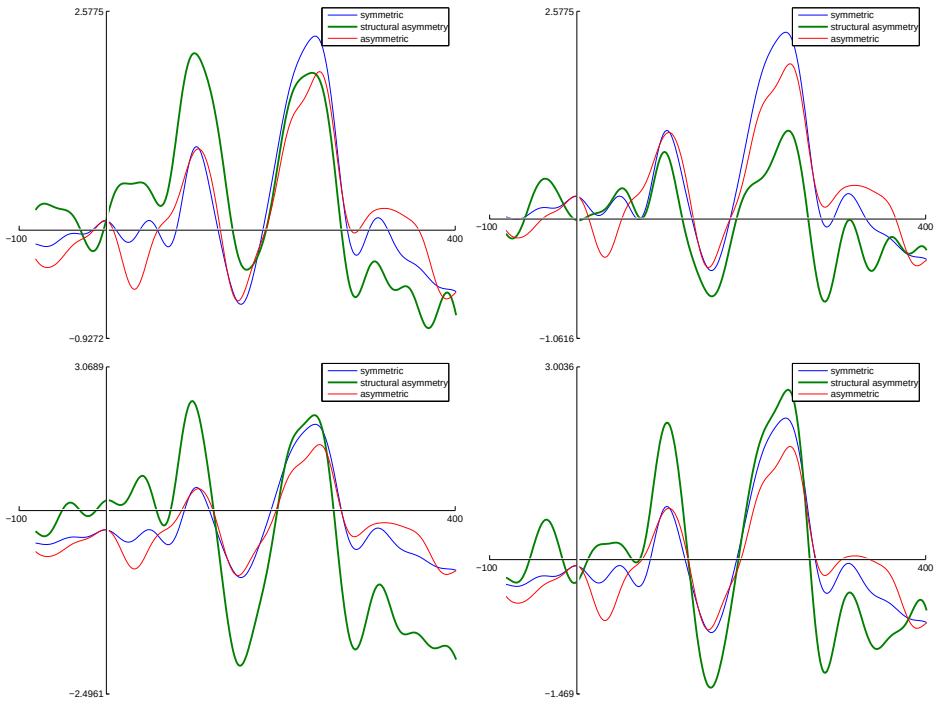

Figure 23: Po1 ERP differences of qualitative asymmetries against symmetry and asymmetry conditions: correct (top) and incorrect (bottom) responses. Both small differences (left) and large differences (right) suggest P2 activation indicates perceived symmetry.

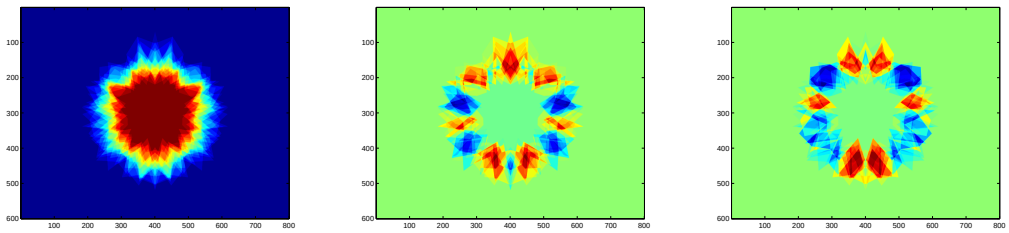

Figure 24: First three leading principal components for quantitative small asymmetry condition. The contrast region (center, right) represents the range of the stimulus contours.

higher amplitude indicates symmetry response. In the case of incorrect responses, several channels show no correlation to either symmetry or asymmetry. It is as if no particular stimulus detection takes place. In either case, P2 seems to indicate the perceived symmetry: the response that was made.

5.4.1 Early symmetry processing

The results in Locher and Nodine (73) and Palmer and Hemenway (86) motivate a two-stage model of symmetry assessment (22). The first stage serves as a fast assessment and may rely on the magnocellular channel of the visual system. The second stage is slower, using direct feature comparisons. ERP differences shown here seem to indicate differences in the former stage. Given these results, can we characterize the first stage further?

If we examine the information content of the stimuli, we might see what type of information is being used in early symmetry assessment. We perform a principal components analysis of the stimuli, and note features within the leading components. One interpretation of these components are that they explain the variance of the distribution of information in the images. The leading eigenimage in this case shows the leading feature explaining variance in these cases is the contrast between the white background and the filled center region of the polygons common to all the stimuli. The second and subsequent components highlight the region of the contours of the stimuli. The result is roughly a concentric range where contour differences exist (see Figure 24).

How does this account for performance differences within small conditions of structural asymmetries? In the case of small quantitative differences, insufficient information is present

for a fast assessment of the contour. In the case of small qualitative differences, more contour information is available in the additional vertex difference; the result is improved performance. In the case of large differences, sufficient contour information is available for fast assesment; the result is relatively high performance accuracy. These patterns suggest that a process like Principal Components Analysis (PCA) (106) could be used by the visual system to assess contours.

CHAPTER VI

SPARSE REGULARIZATION ICA IN ERP ANALYSIS

An overarching objective tying together both the feature selection and Psychophysiological work presented here is demonstrating the usability of mental rotation as a task in a brain-machine interface. We have already seen the relative speed ERPs are captured: on the millisecond scale. Identifying task relevant changes at this time resolution has the potential of creating a robust, effective interface. However, from a classification perspective a number of challenges remain for handling ERP data. the ideal classifier would be able to classify a single ERP. In practice, several trials are recorded and averaged in an attempt to produce more consistent features for classification. Averaging even a few trials improves classification by eliminating a number of the issues related to ERP signals shown in chapters 4 and 5, namely noise.

Even for the most easily detected ERP components, single-trial classification of ERPs is difficult as a result of exogenous noise elements and the relative instability of the signals themselves. Being able to perform classification of ERP data with high accuracy represents a significant contribution to BMI research and a significant step in applying any type of dissociable ERP to BMI control tasks. The benefits of real-time single trial classification of ERP signals extend beyond BMI. For instance, real-time stimulus delivery based on accurate classification of current brain state enables experimental designs that were previously not possible. In terms of BMI, the long term goal is to be able to identify different intentful patterns of activity given a stream of EEG data.

ERPs are presently a fast means of observing user intent, but these signals typically contain a high degree of noise and are generally not stable between trials (*i.e.* time-domain representations are nonstationary due to random perturbations of the signals). Frequency space representations also do not show a sufficient degree of inter-trial coherence for classification. Typical approaches to ERP analysis often involve applying transformations like

ICA to the time-domain representation of the signals (83). In this set of experiments, we examine the relative performance gains in classification using raw, ICA, and sparse regularized ICA representations of ERP data from the two sets of the experiments presented in chapters 4 and 5. The results presented here show that the classification performance for contrasting rotation tasks (according to rotation angle) is sufficient for use in a BMI control loop where a directional component is needed.

6.1 Introduction

Improving classification accuracy of single-trial ERP data according to mental task directly affects ERP based BMI efficiency: fewer trials needed to correctly classify user intent increases the throughput of a BMI. The objective of this work is to increase accuracy of classification of ERP data. ERPs of the same phenomena seem to vary widely as a result of the various noise factors related to EEG. We propose using sparse regularization independent components analysis to improve interpretability of recovered sources. We demonstrate this relative improvement by showing improved classification accuracy using sparse regularization source separation over source separation alone and raw data. We use these features for classification using a Support Vector Machine.

We have demonstrated that under certain conditions, sparse regularization source separation improves the interpretability of recovered sources in noisy and overcomplete cases. ERP data typically contains these elements. Our expectation is that sparse regularization will indeed improve interpretability of recovered sources, and assist in classification of ERP waveforms. Here, we demonstrate the relative improved source recovery in the two ERP datasets for mental rotation and symmetry.

6.2 Previous work

Single-trial classification of ERP data has long been an objective for both BMI and Psychophysiology researchers. Single-trial classification is the ability to correctly classify a single trial sequence of ERP data. The standard analysis techniques for ERP is to use *signal averaging* over trials to obtain stable waveforms that correspond to perceptual and task-related components. Averaging addresses issues of scaling, temporal shift, and noise

factors that all contribute to inter-ERP differences. Single-trial classification takes the alternative objective of classifying each ERP separately. Application of single-trial classification methods assists integration of fMRI and EEG imaging methods (37).

6.2.1 General approaches

We consider previous approaches according to some desiderata: general approaches to classifying large ERP segments; approaches that have been targeted at shorter segments or specific ERP components, typically lasting 50-100ms; and transformational approaches that perform a transformation on time-domain features, chiefly ICA. General learning methods have been applied to large single-trial ERP segments. Typical supervised classifiers have been used to test proposed methods, including support vector machines using the standard nonlinear kernel functions (27; 90; 61). Again, a key issue for performance of these methods lies in the time series stationarity of ERP components. Where the relative stationarity of components does not correlate, supervised methods are poor predictors of performance. Adlakha (2) applied a Support Vector Machine to ERP data, demonstrating the need for stationary ERP components for successful classification. Hidden Markov model approaches have been used to address nonstationary components (84; 88).

6.2.2 Early ERP components

Other graphical models have been applied to the classification task. Kalman filter methods perform well for smoothing data as a preprocessing step to classification (80; 112). This strategy is similar to the method we propose: to employ feature selection methods like ICA as a preprocessing step to classification. Another strategy is to use the spectral properties of the ERP (115; 103; 95). For early ERP components, spectral methods may not detect relevant features in frequency space. Another approach has been to use domain knowledge of the ERP or the specific target components to improve classification. In these cases, the target component is known *a priori* and properties of the component inform the classification (49; 6; 101). P300 response data is an important example where these methods show success. The lateralized readiness potential (LRP) or Bereitschaftspotential (BP) is another

ERP phenomenon where domain knowledge of the target component of the waveform assists classification (11). Kohlmorgen and Blankertz (68) used the waveforms obtained from averaging in a generative model of the ERP signal. In contrast, other approaches focus on learning methods themselves, rather than specific properties of the data (120; 121).

6.2.3 Performance measures

Our objective is to show the applicability of mental rotation in BMI applications. In this work we present both accuracy and throughput performance of our classifier methods and compare to previous results. In traditional supervised learning research, classifier performance is usually depicted as overall accuracy. Interface results are presented as throughput or bitrate. McFarland et al. (77) defines bitrate as:

$$B = \log_2 N + P \log_2 P + (1 - P) \log_2 [\frac{1 - P}{N - 1}] \tag{19}$$

where B is the bit rate, N is the number of targets, and P is the probability of success. Dividing B by the length of the trial in minutes computes the bits/min. In comparing classification accuracy and bitrates of our methods to other published results, we demonstrate its effectiveness as a usable interface.

Of course, an important contribution is the interface itself; Randolph (96) demonstrates that not all BCI systems work equally effectively for all users. The disparity is due to the fact that different Psychophysical responses are easier to detect in some people than others. The interface presented here represents a novel contribution as another possible alternative for users where other interfaces do not work well.

6.2.4 Transformational approaches

Previous methods have used ICA as feature selection methods for classification (113; 94; 83). The intuition for using ICA features is that while the relative ERP components change as a result of noise and other factors, the information of the perceptual components of the ERP are preserved between trials. Ashtiyani et al. (4) used ICA feature selection as features for neural network classification. Potter et al. (94) used ICA features in a support vector machine classifier. Oveisi (85) developed a nonlinear ICA method and shows relative

performance gains in ERP data.

6.3 Objective

The goal of single-trial classification also provides an important real-world problem that allows us to test the relative performance gains of sparse regularization ICA against conventional ICA methods, particularly infomax ICA. There are several empirical objectives for these experiments. First, the classification accuracy of the feature sets between conventional ICA, sparse regularization ICA, and raw data allows us to compare the relative effectiveness of these methods on actual ERP study data. An important objective of these experiments is to show the relative performance gains using rICAs feature selection over raw data and ICA (infomax) generated features. While the long-term goals of this work are to maximize overall accuracy rates for BMI, the immediate goal is to show that rICAs generated features improve classification accuracy, and therefore produce features that are better representations of the functional processes behind the ERP data.

We also compare the relative effectiveness of the two types of ERP data features collected in the respective studies. In the case of the symmetry ERP study, the target ERP components are early (before 300ms) while in the rotation studies the latter components are the target (after 300ms). Finally, these experiments also indicate the viability of these effects for use in a brain-machine interface. Being able to classify differences in these phenomena with accuracy and bitrates comparable to the state-of-the-art provides evidence of their viability in a brain-machine interface.

6.4 Experiments

These experiments consist of labeling trial data, sampling a training and testing subset of the trial classes, and applying classification methods to the subsets. In the experiments presented here, we use a support vector machine classifier that uses a radial basis function kernel (15). We apply n-fold cross validation at training time to learn optimal values for the C and gamma parameters for a radial basis function (RBF) kernel. The datasets are taken from the rotation and symmetry studies presented in earlier chapters. Recall in both experiments that six practice trials were presented to participants as initial training for the

task; the data from these practice trials were not used in any analysis. Six practice trials is a common practice in ERP studies (31). These practice trials serve to make participants comfortable with the performance task and to lower learning effects.

We select a subset of trials for training and testing from each participant dataset, selecting equal numbers of trials of each condition from each participant. In both datasets we draw training and testing samples from early and latter trials respectively. For each subject, training data is taken from early trials. For the symmetry dataset, the first 100 symmetry and total asymmetry trials are the training data; the next 100 symmetry and total asymmetry trials are the test data. For the rotation data, the first 20 trials of the four angle bins (left, right, up, down) are used for training; the next 20 are used for testing. Instances were selected sequentially as they would be required in actual use: a training set is gathered for the classifier before the interaction session begins.

6.4.1 Feature generation/classification procedure

We use an experimental setup similar to the SVM strategy suggested by Adlakha (2). We first used the raw features of the ERP data collected from the two studies in the support vector machine classifier. We rasterized the time series data of two of the EEG channels in each experiment to form a feature vector for each instance (trial): for symmetry, P3 and T6; for language, Ft7 and P4; and for rotation, P3 and P4 in the 10-20 system (see Section 4.7.2). Data were mean corrected and rectified. We applied the classifier to the training dataset (early trials) and tested performance using the test set (latter trials). In both the ICA and srICA feature cases, we applied the source separation method to each trial. We then rasterized the recovered features to form an instance (trial) feature vector. We trained the classifier on the generated training data. For the test dataset, we again applied source separation to each trial in the test set for the ICA and srICA experiments, collecting instances in the same manner as training to form the instance matrix. We tested the trained classifier performance on the recovered test sets.

6.4.2 Symmetry data

Recall from the symmetry ERP experiment (Chapter 5), a significant difference in the early P2 ERP component between the symmetry and asymmetry conditions. These are the cases where symmetry is exact and where total asymmetry are shown to participants. Participant accuracy rates are highest in these two conditions. We hypothesize the samples in the time series of this component are best for classifying ERPs between symmetry and asymmetry conditions.

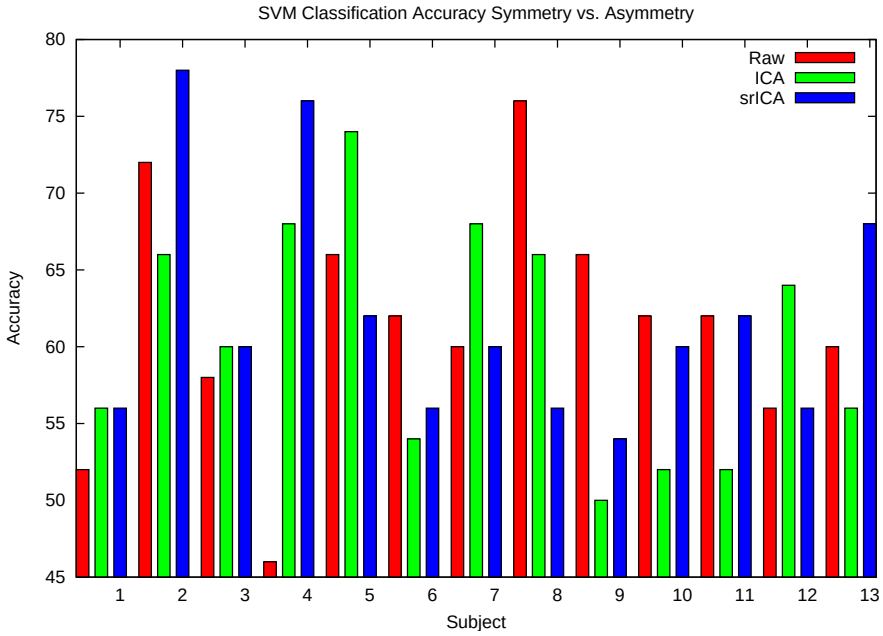

Figure 25: Symmetry vs. Asymmetry conditions classification accuracy per subject.

For the symmetry versus asymmetry conditions in the symmetry ERP study dataset, rICAs accuracy was higher than raw features for 62% of subjects; rICAs accuracy was higher than ICA features for 69% of subjects. However, the differences between the three methods was not statistically significant. Further, overall accuracy across subjects for all three methods was about 60% (0.613, 0.604, and 0.618 for raw, ICA and rICAs features respectively; see Figure 25).

6.4.2.1 Language production versus rotation

We first assess classification accuracy between the language production ERP and rotation. In this experiment, many trials were rejected on the basis of low signal to noise ratios. Trials were visually inspected for artifact content and rejected manually. After trial rejection, four subjects contained the minimum number of trials for training and test sets. The classification performance was comparable to performance on the symmetry dataset. As in the symmetry dataset, there was no significant difference between the feature selection methods (see Figure 26).

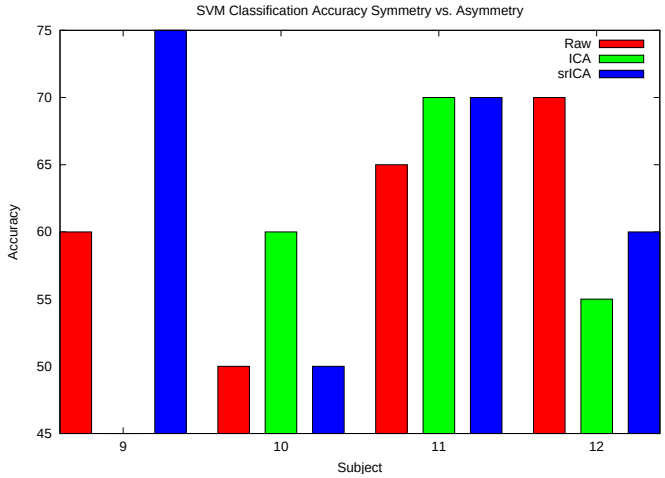

Figure 26: Rotation vs. language classification accuracy.

6.4.2.2 Mirror rotation angles

To assess classification accuracy for rotation angles, we collect trials according to ranges of angles from the trial data and perform binary classification on the pairs of angle bins (see Figure 27). Trials were collected according to the angle bin and assigned a class label. We applied SVM classification to each angle bin pair (see Figures 28 & 29).

6.5 Discussion

The use of SVM classification, while not directly modeling potential nonstationarity in the data produces a higher overall accuracy rate using srICA generated features over ICA or raw features in the symmetry dataset. While sparse regularization features produces

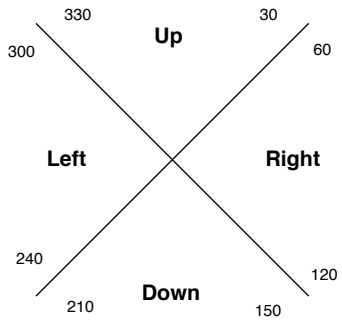

Figure 27: Angle bins for classification.

higher accuracy in a majority of participants' data, the difference between raw data, ICA, and rICAs is not statistically significant. Results were similar for differentiating language production and rotation tasks. In the rotation data experiments, performance accuracy for a majority of subjects improves using rICAs generated features. The relative variability of the results over all subjects requires additional analysis. In this case, we examine the relative accuracy of differentiating rotation activity of a left or right rotation task.

We first demonstrate the generalizability of the experiment by analyzing data gathered from an additional seven subjects in the explicit and covert rotation experiment. We use the same acquisition and analysis paradigms for these data. Figure 30 shows the per subject results of these data for the left/right and up/left and up/right pairwise comparisons.

Again, both sparse regularization ICA and ICA methods generate features that improve classification accuracy for these experiments. This additional experiment demonstrates the relative performance differences are consistent over two sets of participants. We consider this dataset in more detail. We compare classification accuracy in the *unforced choice* setting between left and right trials. Here, we rank test instances by their probability in the opposite class distribution learned from training. We sample without replacement half the testset using the ranking provided by class 0; we sample the other half from the class 1 ranking. Figure 31 shows the mean accuracy of subject data for increasing levels of retention. Mean accuracy is highest where the retained data is smallest. By retaining instances with the lowest probability of the opposite class, the learned classification function's accuracy is maximal. Highest mean accuracy (92%) is obtained at 20% data retention using ICA

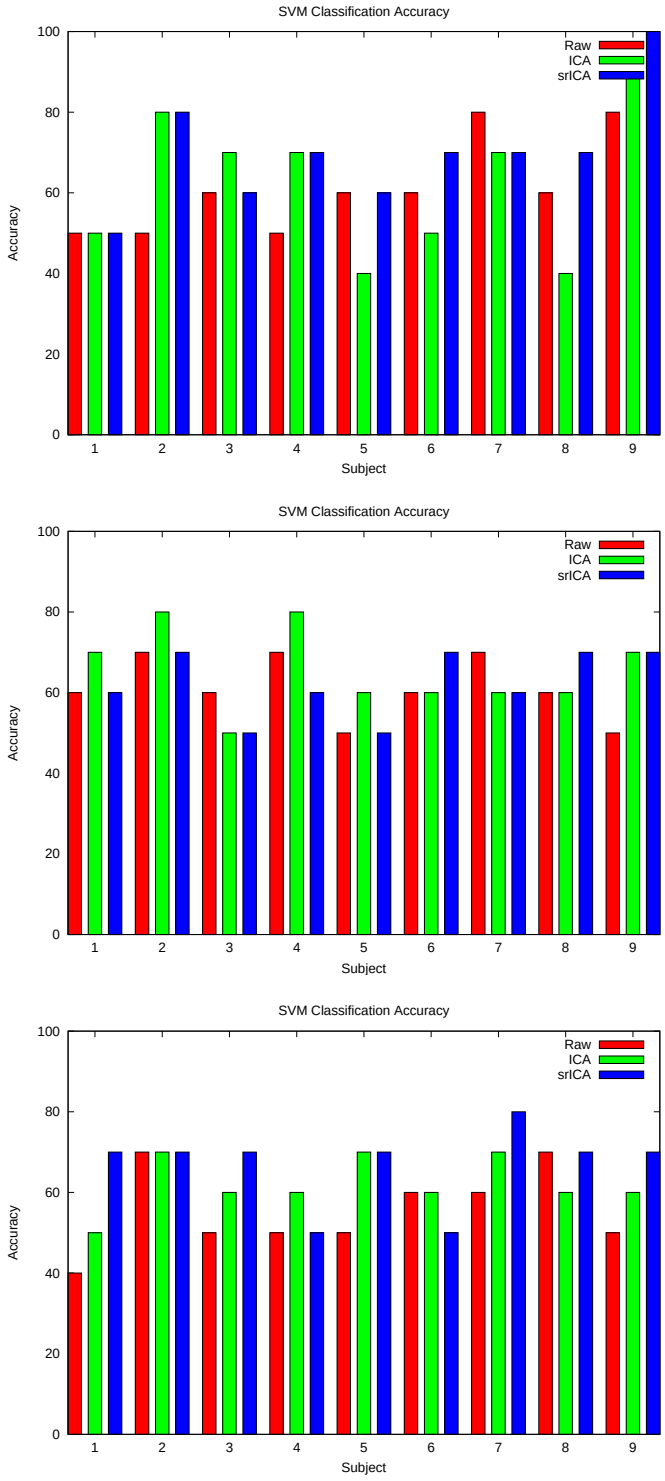

Figure 28: Classification accuracy comparison for left/up,right/up,down/up directions (figures top,middle,bottom respectively).

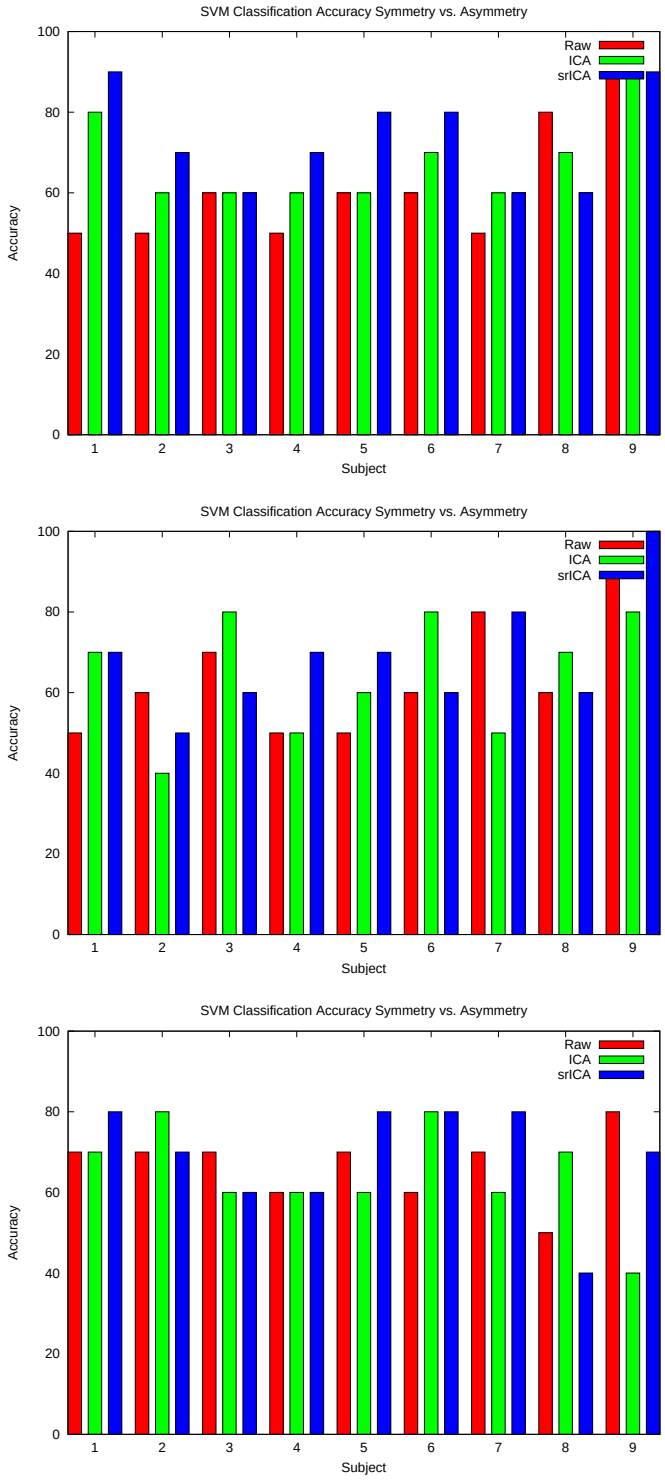

Figure 29: Classification accuracy comparison for left/right, right/down and left/down directions (figures top,middle,bottom respectively).

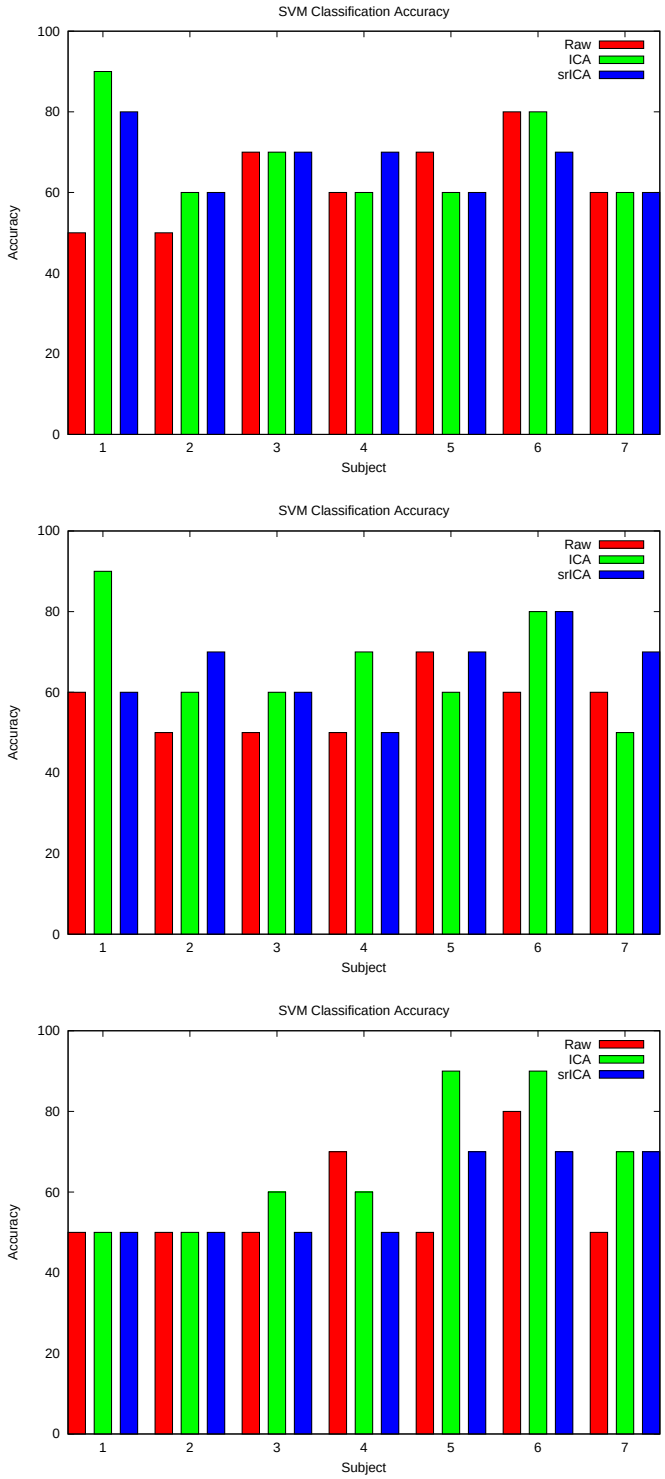

Figure 30: Classification accuracy comparison for left/right, left/up and right/up directions.

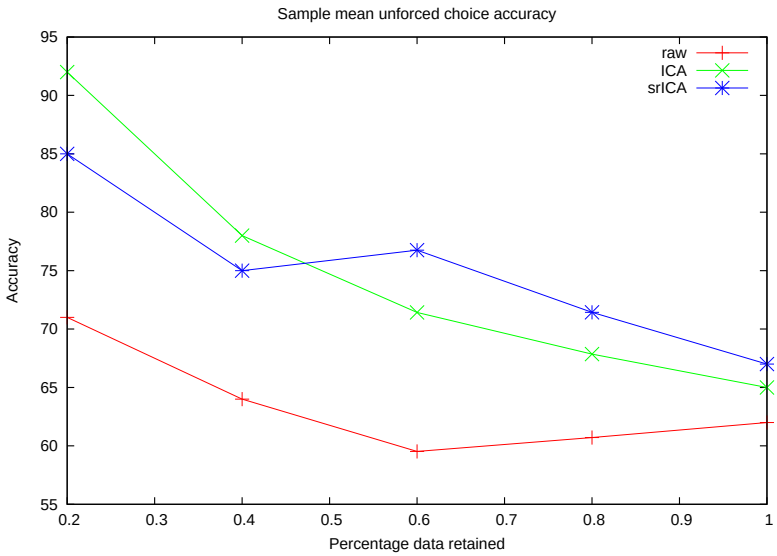

Figure 31: Unforced choice classification accuracy for raw, ICA, and srICA feature selection.

features; lowest mean accuracy (59.5%) is obtained at 60% data retention using raw features. In terms of overall performance, raw features perform worst at every data retention level; srICA comparable or slightly better than ICA at latter levels.

The bitrate for the srICA method where 60% data are retained is 11.1 (see Table 4). The bitrate for the srICA method where 100% data are retained is 7.231. At 80% data retention using srICA feature selection, the bitrate is comparable to the bitrate comparison results presented by McFarland et al. (77) and the accuracy results presented by Cheng et al. (16). McFarland et al. (77) show an averaged bitrate of 8.385 for optimal performance in a target selection task. In this case, the user task was to direct a cursor to a one of several target locations. Each subject attempted a set of trials for an increasing number of targets. The highest bitrate was considered over all the target numbers. The sample mean bitrate for the two target task performance used by McFarland et al. (77) is 5.724. Comparing our results to these two results is relevant to demonstrating relative usability of mental rotation: the cursor/target task (RJB) used by Cheng et al. (16) and McFarland et al. (77) is an included interface task in BCI2000 (99).

A large factor in the bitrate for mental rotation is the relative number of attempts per unit time; being able to run many trials per minute increases the bitrate, despite a relatively

Table 4: Classification accuracy and bitrates for averaged unforced choice retention levels.

Method	Retention	Accuracy	Bitrate
raw	0.2	0.71	2.2317
raw	0.4	0.64	1.9488
raw	0.6	0.5953	1.3447
raw	0.8	0.6071	2.2681
raw	1.0	0.62	3.5664
ICA	0.2	0.92	10.163
ICA	0.4	0.78	8.1543
ICA	0.6	0.7142	6.9751
ICA	0.8	0.6786	6.3990
ICA	1.0	0.65	5.6042
srICA	0.2	0.85	6.6327
srICA	0.4	0.75	6.4165
srICA	0.6	0.7675	11.100
srICA	0.8	0.7143	9.3091
srICA	1.0	0.67	7.2313

low accuracy (700ms trial time, under the trial independence assumption). Further, Cheng et al. (16) present performance of three subjects using mu rhythm activity again to guide a cursor to a target. Cheng et al. present classification accuracy (without bitrate) for a novel feature selection method, where best average performance ranged for parameterizations of the method from 69% to 76%.

Tables 5, 6 and 7 show per subject performance in accuracy and bitrates for the three feature selection settings (raw,ICA,srICA respectively). These tables show in detail the accuracy to bitrate comparison for each subject for increasing levels of data retention in the unforced choice paradigm.

Table 5: Per subject unforced choice accuracies/bitrates for raw features. Each pair of acc,bitrate columns lists performance for retention levels (0.2,0.4,0.6,0.8,1.0 respectively).

subject	acc	bitrate	acc	bitrate	acc	bitrate	acc	bitrate	acc	bitrate
1	.50	0.	.50	0.	.50	0.	.50	0.	.50	0.
2	.50	0.	.50	0.	.50	0.	.50	0.	.50	0.
3	1.00	17.	.75	6.4165	.6667	4.1686	.625	3.0985	.70	10.090
4	.50	0.	.75	6.4165	.6667	4.1686	.625	3.0985	.60	2.4692
5	.50	0.	.50	0.	.50	0.	.625	3.0985	.70	10.090
6	1.00	17.	.75	6.4165	.6667	4.1686	.75	12.833	.80	23.636
7	1.00	17.	.75	6.4165	.6667	4.1686	.625	3.0985	.60	2.4692

The per subject results show additional details of the usability of mental rotation. We

Table 6: Per subject unforced choice accuracies/bitrates for ICA features. Each pair of acc,bitrate columns lists performance for retention levels (0.2,0.4,0.6,0.8,1.0 respectively).

subject	acc	bitrate	acc	bitrate	acc	bitrate	acc	bitrate	acc	bitrate
1	.5	0.	.75	6.4165	.6667	4.1686	.75	12.833	.7	10.090
2	1.	17.	.75	6.4165	.6667	4.1686	.625	3.0985	.6	2.4692
3	1.	17.	1.	34.	.8333	17.845	.625	3.0985	.6	2.4692
4	1.	17.	.75	6.4165	.6667	4.1686	.75	12.833	.7	10.090
5	1.	17.	.75	6.4165	.6667	4.1686	.625	3.0985	.6	2.4692
6	1.	17.	.75	6.4165	.6667	4.1686	.625	3.0985	.6	2.4692
7	1.	17.	.75	6.4165	.8333	17.845	.75	12.833	.8	23.636

Table 7: Per subject unforced choice accuracies/bitrates for srICA features. Each pair of acc,bitrate columns lists performance for retention levels (0.2,0.4,0.6,0.8,1.0 respectively).

subject	acc	bitrate	acc	bitrate	acc	bitrate	acc	bitrate	acc	bitrate
1	.5	0.	.5	0.	.6667	4.1686	.75	12.833	.6	2.4692
2	1.	17.	.75	6.4165	.833	17.809	.625	3.0985	.7	10.090
3	1.	17.	1.	34.	1.	51.	.875	31.038	.7	10.090
4	1.	17.	.75	6.4165	.6667	4.1686	.75	12.833	.7	10.090
5	1.	17.	.75	6.4165	.6667	4.1686	.625	3.0985	.6	2.4692
6	.5	0.	.75	6.4165	.6667	4.1686	.625	3.0985	.6	2.4692
7	1.	17.	.75	6.4165	.833	17.809	.75	12.833	.8	23.636

consider subject performance using srICA features (Table 7). First, subjects 1,5, and 6 all perform poorly at almost all levels of data retention in the unforced choice classification. In contrast, subjects 2, 4, and 7 show consistent performance at almost all levels. Finally subject 3 performs well at all levels, performing at ceiling or better than all subjects in four of five retention levels. In this case, unforced choice classification identifies those subjects who perform the task well. This corroborates the evidence gathered by Randolph (96): performance differences in Psychophysical tasks leads to interfaces that are better suited to particular populations.

Being able to recognize both left and right rotations when compared with non-task related activity provides further evidence of differentiating mental rotation tasks from other intentful activity in the brain. Admitting unforced choice allows us to consider the case where we exclude trials that resist accurate classification. Bitrates that are comparable to widely used interfaces demonstrate the relative usability of mental rotation in BMI.

CHAPTER VII

CONCLUSION

7.1 Conclusions

In this dissertation, we have presented a set of methods aimed at improving the discriminable power of time-varying features. We have approached the problem from both a dimensionality reduction as well as a regularization perspective and have shown significant improvement in noisy, overcomplete situations. In the case of dimensionality reduction, we have shown decreased type II errors in actual fMRI study data. In the case of regularization, we have shown improvement in noisy and overcomplete cases in simulated data. The implications of this work are that regularization is a powerful method for factoring noise from data in the context of source separation.

We have also presented results of Pyschophysiological studies measuring ERP differences to visual stimuli. The two studies focus on two types of differences in ERPs: temporally early component differences and temporally late component differences. In the former case, the results of the study show early ERP activity correlated with symmetry assessment. In the latter case, the results of the study show differentiable differences of the ERPs between language production and mental rotation as well as between angular conditions of mental rotation tasks themselves. Finally, we have shown evidence of the effectiveness of using mental rotation tasks in a BMI. By using unforced choice classification, we demonstrate that sparse regularization ICA improves classification accuracy. The classification accuracy and bitrates in a typical use case for a BMI using mental rotation are comparable to results found using popular interfaces in BMI research. Mental rotation tasks represent a novel alternative Psychophysiological phenomenon for control. The use of sparse regularization as feature selection in mental rotation produces classification accuracy and bitrates that are comparable to popular inteface methods for BMIs. The results show the potential of using mental rotation as the mental task in a BMI.

7.2 Future Work

The methods and experiments presented in this dissertation represent a beginning point for a number of future explorations in the areas of feature selection, ERP analysis, mental rotation, and brain-machine interfaces. First, the space of feature selection methods that use regularization strategies has not been fully explored. The long-term objective of this work is to find methods that produce globally optimal methods for feature selection in the context of regularization for noise models. In particular, embedded regularization within feature selection methods create efficient means of combining the two objectives.

Second, there is a growing need for intelligent tools for analysis of ERP data. As analytical demands increase as a result of more sophisticated experimental design, the need for intelligent tools increases. Integration of information from many types of sensors is yielding additional information from imaging data (46). Applying source separation methods that simultaneously reduce noise or reduce dimensionality represents an important step for research in multi-modal data.

While the experiments on mental rotation presented in this dissertation begin to address current Psychophysiological debates over mental rotation activity, a great deal of work still remains in determining mental rotation's participation in the dorsal visual pathway. Imaging mental rotation tasks using fMRI should reveal the subtle differences in activation noted here in ERP experiments. More resolute ERP data of mental rotation should also reveal more details of activation differences. Focusing additional sensors over parietal regions of the brain is another means of obtaining clearer data of mental rotation tasks.

Finally, the results presented in this dissertation of differentiating mental rotation tasks for BMI represent only a start to realizing mental rotation based BMI. Chief among these tasks is implementing an rotation based interface itself. Likely target interface applications are those that have a directional component to them. Cursor to target control (*e.g.* mouse pointer control) is one application often used in BMI results. Typically, these are constrained, where only one direction is under BMI control (77). A significant contribution would be the use of mental rotation to control both dimensions in a two-dimensional control

task. Another potential benefit of mental rotation in BMI addresses the Midas touch effect (110). A set of experiments aimed at showing the acute modulation of mental rotation over time represents a significant, novel contribution to BMI research.

APPENDIX A

FACTORING SLOW NONLINEAR TRENDS IN ERP DATA USING SPLINE REGRESSION

A.1 Slow nonlinear artifact removal

Digital filtering is a standard practice in removing artifacts from noisy sensor data like electroencephalographic data. These filters are able to capture many of the common exogenous artifacts present in EEG data. A common design uses Butterworth filters to remove artifacts outside the frequency response spectra of EEG data; artifacts from electrical devices and outlets (approximately 50-60Hz) are well accounted for using these filters. Other artifacts, however are not as easily captured using these filters. Artifacts that overlap the frequency response of target ERP signals and nonlinear trends are not well captured using these filters. In active electrode arrays, participant perspiration often can introduce a nonlinear trend to data, and often results in trial rejection. In this work, we introduce a method based on smoothing splines that factors nonlinear trends while minimizing effects to target frequency response ranges.

Artifact removal is a common preprocessing step in analysis of many types of sensor array data and often affects performance of subsequent analysis methods. Electroencephalography (EEG) data is a representative of a class of sensors where there are several types of noise present that affect target signal quality. Signals caused by processes other than the target signal are almost always present in the raw data and are often several orders of magnitude larger than the signals of interest. We seek a method to remove low-frequency nonlinear trends from data that deviate from baseline activity while preserving target signal.

EEG signals contain noise from a number of independent factors that are based on the physics of electrode sensing of electrical potentials. Artifacts from the environment consist of RF emissions of seemingly ubiquitous electrical devices as well as electrical outlets themselves. Within the brain, target electrical potentials are mixed with potentials from

many cranial regions that activate for non task related purposes. Conductance properties of the cranial tissue transmit potential in unpredictable ways. Electrodes themselves react with the conductive properties of the scalp introducing artifacts from perspiration and other epidural dynamics. Artifact signals caused by the potentials emitted from muscle movements often dominate target signal emissions. Muscles proximate to recording sites emit high amplitude potentials relative to target signals. Worse, they share many of the frequency characteristics of target signals. In practice, trials are rejected based on the presence of dominating muscle noise. Methods for removing muscle noise would benefit experimental researchers, recovering trials previously lost to muscle artifacts. Together, these artifacts produce a challenge for discovering target signals in EEG data. Indeed, many researchers opt for more invasive sensors to reduce the effects of these artifacts.

First, we consider previous work in EEG artifact removal. A common practice is filter design motivated by traditional signal processing methods. These methods are successful in removing artifacts where the frequency characteristics between target and artifact are sufficiently different so that discrete frequency bands are selected to optimize target signal segmentation.

Independent components analysis is a popular method for removing artifacts, where independent components of artifacts are identified and factored from analysis (62; 67). This typically requires expert intervention in identifying artifact components as the waveform for many artifacts is nonlinear and non-stationary. Component methods are used for EOG blink removal in part because the waveform of the blink is well known. In non-square cases, the components may represent mixtures of components that are independent of the other components. In other words, one is not always guaranteed ideal separation of artifact signals. Other artifact signals may have Gaussian distributions, the presence of many of these signals may pose a challenge for source separation methods.

As a large class of artifacts that affect many types of recording paradigms, EOG has received a large amount of attention (51; 57; 65; 78). Gratton et al. (47) proposed a now popular rejection method for removing EOG artifacts from EEG data. EOG signals of eye muscle movements are represented in EEG recording as relatively short duration, high

frequency signals

Spline methods take advantage of the piecewise nature of many functions by connecting a set of lower degree functions at fixed points. The intuition is that the intervals between fixed points are well modelled using lower degree functions. Splines were developed to represent free-form curves and surfaces, and have broad application in graphic rendering methods for free-form objects. In general, piecewise methods are attractive for fitting functions to data, where local regularity is related to smoothness properties. In this case, we use piecewise methods to fit segments of the nonlinear trend in the complete dataset and remove them for further analysis. Smoothing splines also have important regression properties that we apply to artifact removal. We also present an extension to smoothing spline knot selection

A.1.1 Cubic smoothing splines

Cubic smoothing splines are common tools for regression where smoothness of the regression function is desired. Consider a typical regression formulation, where a set of instances $x_{ij}, i : 1 \dots n$ are represented as a set of predictors or independent variables $j : 1 \dots p$ and dependent variables y_i. The goal of regression methods is to find a set of weights β_j that minimize the error of the predictors to the dependent variables:

$$\underset{\beta}{\operatorname{argmin}} |y_i - \beta x_i|_2^2 \tag{20}$$

In regression for time dependent variables, we consider fitting a function $g(x)$ between pairs of points. In this case, regression for $g(x)$

$$\sum_{i=0}^{n} w_i [y_i - g(x_i)]^2 \tag{21}$$

Finally, where smoothness of the regression line is needed, the continuity property of $g(x)$ is used. Where g is a cubic polynomial, second order continuity has smoothness properties sufficient for most requirements. A parameter α represents the trade-off between the regression fit and smoothness:

$$J_\alpha(g) = \alpha \sum_{i=0}^{n} w_i [y_i - g(x_i)]^2 + (1 - \alpha) \int_{x_0}^{x_n} g^2(x)^2 dx \tag{22}$$

The objective here is to find optimal alpha parameter settings for removing nonlinear artifacts. In terms of EEG and nonlinear trends caused by additive artifacts $s + x$, the goal

is to model artifact signals for subtraction while preserving the information content of the EEG signal. An important observation is that these artifacts share many of the frequency characteristics of the information signal in EEG. Smoothing splines seem to remove the

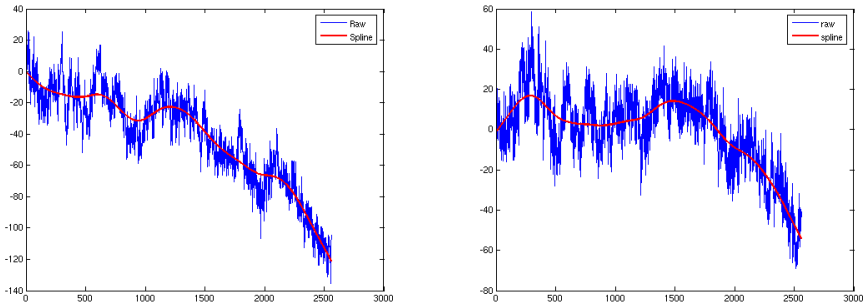

Figure 32: EEG data with nonlinear artifact (blue) with smoothing spline (red). $\alpha = 10^{-9}$

nonlinear trend while preserving information signal in cases where artifact is present, and where signal is relatively artifact-free.

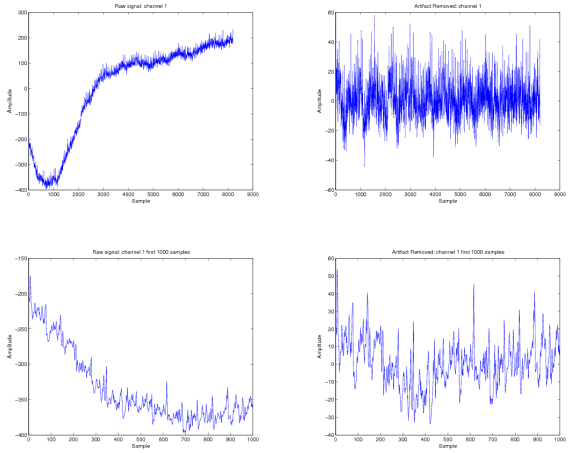

Figure 33: Detail of EEG data before and after artifact removal.

A.2 Experiments

The experiments are aimed at testing the relative performance in recovery of ERP components in data with perspiration artifact using the frequency band specification optimization. Here, we use the same baseline (noise containing and noise free) ERP data. we test performance of the smoothing spline subtraction using a P300 signal inserted in artifact data. In this case, the P300 signal is added to a single channel containing artifact. The

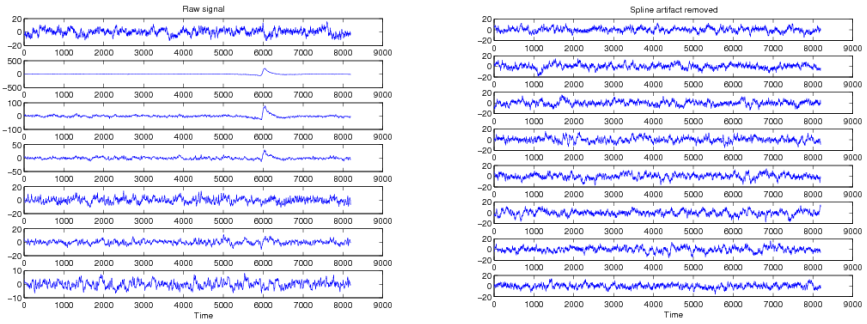

Figure 34: Effects of smoothing splines on artifact free data.

artifact is present over all trials, and the P300 signal is added to the signal starting at the common time-locked event (stimulus onset). Averaging analysis is used on the data with and without application of the smoothing splines. We measure MSE of the recovered P300 signal in both cases. There is a small difference between the two waveforms (MSE=171.88). However, when comparing to the original p300 waveform, the difference is larger: RMSE(spline)=48.39 RMSE(averaged)=50.49 (see Figure 35).

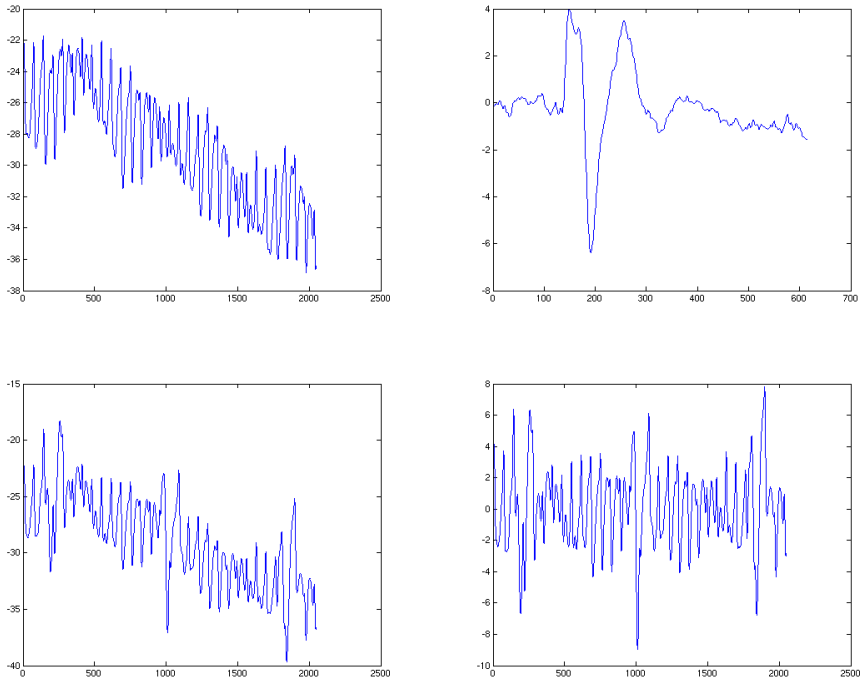

Figure 35: Noisy signal (top, left). Information signal: a P300 signal inserted at regular intervals (top, right). Averaged signal (bottom,left). Averaged signal after spline removal (bottom,right).

A.3 Discussion

We have presented an initial experiment demonstrating the practical viability of smoothing splines to factoring slow nonlinear artifacts in EEG data. The smoothing spline method makes few assumptions of the noise model. In the case where a spline knot is used at each datapoint, the spline can fit the data with zero reconstruction error. An alternative method to factor these noise artifacts would be to use polynomial regression. Polynomial regression fits a polynomial function to data by finding best-fit coefficients to terms of the polynomial function. The method uses a restriction bias and requires a parameter that represents the order of the polynomial used to fit to data. In contrast to polynomial regression methods, smoothing spline regression captures the nonlinear artifacts of perspiration without the model constraints imposed using polynomial regression. In most practical cases, perspiration artifact adds complex nonlinear trends that make prediction of the optimal polynomial order difficult to estimate.

A.4 Learning optimal spline parameters

From a learning perspective, parameterization of the smoothing spline for optimal removal of the artifact while maximally preserving target signal components is typically done using cross validation over the alpha value. In the case where we want to perform real-time artifact correction using splines, cross validation does not seem effective. Further, from the practioner's perspective, the alpha parameter is difficult to reason about. In the context of the smoothing spline equation, an alpha value that is too small does not capture the artifact by placing too much weight on the smoothing term. Alpha values that are too large overfits to the data, removing high frequency information from the signal. The effects of the spline on the target signal data are much harder to predict in terms of the alpha parameter. Ideally, we want to find the alpha that removes the low frequency content while maximally preserving the objective signals, and tuning $0 <= \alpha <= 1$ is not a reasonable search for practioners. Rather, we can reformulate alpha parameter selection in terms of target frequency bands to preserve.

First, we note the Discrete Fourier Transform (DFT) provides the frequency content

information over a discrete range. We can constrain the frequency content of the spline regressor to exclude the part of the frequency range that we want to preserve. In the context of ERP and spectral analysis for EEG data, this is easier to quantify than the alpha parameter setting for the smoothing spline equation. Specifying this frequency range allows us to minimize the 2-norm of this frequency range of the spline regressor itself. We optimize the minimization of this norm over alpha.

The objective function for this optimization that minimizes the sum of the slope of the best-fit linear regressor and the 2-norm over the discrete high-frequency selection:

$$\underset{\alpha}{\operatorname{argmin}}\, L(X - J_\alpha(g))^2 + ||F_{a,b}(X - J_\alpha(g)) - F_{a,b}(X)|| \tag{23}$$

where α represents the spline tradeoff parameter, F represents the discrete Fourier transform, J is the smoothing spline signal parameterized by α, g the (cubic) spline function, (a, b) represents the discrete target frequency range, and X represents the observed signals. We first show the gradient solution to the optimization by expanding the objective function:

$$F(J_\alpha(g), a, b) = \sum_{n=a}^{b} j_n e^{-i2\pi k \frac{n}{N}} = v_k \tag{24}$$

$$F(X, a, b) = \sum_{n=a}^{b} x e^{-i2\pi k \frac{n}{N}} = u_k \tag{25}$$

$$v_k - u_k = \sum_{n=a}^{b} [[\alpha w_n [y_n - g(x_n)]^2 + (1 - \alpha) \int g^2(x_n)^2 dx] e^{-i2\pi k \frac{n}{N}} - x_n e^{-i2\pi k \frac{n}{N}}] \tag{26}$$

The derivative with respect to α leaves the terms of the spline:

$$\sum_{n=a}^{b} [w_n [y_n - g(x_n)]^2 j + \int g^2(x_n)^2 dx] e^{-i2\pi k \frac{n}{N}} \tag{27}$$

The interpretation of this objective is that changes in alpha affect the spline terms of the objective α.

In the case of EEG data for ERP or spectral analysis, this is a practical means for practitioners to specify what signal ranges to preserve. As a matter of practice, the signal

drifts noted in the case of perspiration are low frequency such that the normal discrete filter ranges used in analysis are sufficient for removing perspiration artifact. In other cases, the method is capable of removing frequency content to arbitrary precision.

Bibliography

[1] Brain trauma foundation facts. July, 2009. `http://www.braintrauma.org`.

[2] A. Adlakha. Single trial EEG classification. Technical report, Swiss Federal Institute of Technology, 2002.

[3] B. Alivisatos and M. Petrides. Functional activation of the human brain during mental rotation. *Neuropsychologia*, 35:111–118, 1997.

[4] M. Ashtiyani, S. Asadi, P. Birgani, and E. Khordechi. EEG classification using neural networks and independent component analysis. In *Proceedings of the Fourth Kuala Lumpur International Conference on Biomedical Engineering*, 2008.

[5] M. Babaie-Zadeh, C. Jutten, and A. Mansour. Sparse ICA via cluster-wide PCA. *Neurocomputing*, 69:1458–1466, 2006.

[6] C. Bandt, M. Weymar, D. Samaga, and A. Hamm. A simple classification tool for single-trial analysis of erp components. *Pyschophysiology*, 46:747–757, 2009.

[7] J. Barnes, R. Howard, C. Senior, M. Brammer, E. Bullmore, A. Simmons, P. Woodruff, and A. David. Cortical activity during rotational and linear transformations. *Neuropsychologia*, 38:1148–1156, January 2000.

[8] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, June 2003.

[9] A. J. Bell and T. J. Sejnowski. An information maximization approach to blind separation and blind deconvolution to blind source separation and blind deconvolution. *Neural Computation*, 7(6)(6):1129–1159, 1995.

[10] Niels Birbaumer, A. Kubler, N. Ghanayim, T. Hintenberger, J. Perelmouter, J. Kaiser, I. Iverson, B. Kotchoubey, N. Neumann, and H. Flor. The thought translation device (ttd) for completely paralyzed patients. *IEEE Transactions on Rehabilitation Engineering*, 8:190–192, 2000.

[11] B. Blankertz, G. Dornhenge, C. Schafer, R. Krepki, J. Kohlmorgen, K.-R. Muller, V. Kunzmann, F. Losch, and Curio G. Boosting bit rates and error detection for the classification of fast-paced motor commands based on single-trial eeg analysis. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 11:127–131, 2003.

[12] P. Bradley and O. Mangasarian. Feature selection via concave minimization and support vector machines. In *Proceedings of the 13th international conference on machine learning*, 1998.

[13] A. Bronstein, M. Bronstein, M. Zibulevsky, and Y. Zeevi. Sparse ICA for blind separation of transmitted and reflected images. *International Journal of Imaging Science and Technology*, 15:84–91, 2005.

[14] H. Buchner, R. Aichner, and W. Kellermann. Blind source separation for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity. In *Proceedings*

of the Seventh International Workshop on Acoustic Echo and Noise Control, 2003. URL `citeseer.ist.psu.edu/article/buchner03blind.html`.

[15] C. Chang and C. Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/ cjlin/libsvm`.

[16] M. Cheng, W. Jia, X. Gao, S. Gao, and F. Yang. Mu rhythm-based cursor control: an offline analysis. *Clinical Neurophysiology*, 115:745–751, 2004.

[17] A. Cichocki, S. C. Douglas, and S. Amari. Robust techniques for independent component analysis (ica) with noisy data. *Neurocomputing*, 22:1423, November 2000. URL `citeseer.nj.nec.com/cruces-alvarez00iterative.html`.

[18] M. S. Cohen, S. M. Kosslyn, H. C. Breiter, G. J. DiGirolamo, W.L. Thompson, A. J. Anderson, S.Y. Bookheimer, B.R. Rosen, and J.W. Belliveau. Changes in cortical activity during mental rotation: a mapping study using function magnetic resonance imaging. *Brain*, 119(1):89–100, 1996.

[19] Ronald R. Coifman and Stephane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.

[20] L. A. Cooper. Mental rotation of two-dimensional shapes. *Cognitive Psychology*, 7: 20–43, 1975.

[21] L. A. Cooper and Roger Shepard. *Visual information processing*, chapter Chronometric studies of the rotation of mental images, pages 75–176. Academic Press, 1973.

[22] M. Corballis and I. Beale. *The psychology of right and left*. Erlbaum and Associates, 1976.

[23] M. Corballis and A. Blackman. The effect of apparant motion on mental rotation. *Memory & Cognition*, 18:551–555, 1990.

[24] M. Corballis and H. McMaster. The role of stimulus-response compatability and mental rotation in mirror-image and left-right decisions. *Canadian Journal of Psychology*, 50:397–401, 1996.

[25] Michael C. Corballis. Mental rotation and the right hemisphere. *Brain and Language*, 57(1):100–121, March 1997.

[26] Alexandre D'Aspremont, Laurent El Ghaoui, M. Jordan, and Gert R. G. Lanckriet. A direct formulation for sparse pca using semidefinite programming. In *NIPS*, 2004.

[27] Stefan Debener, M. Ullsperger, M. Siegel, and A. Engel. Towards single-trial analysis in cognitive brain research. *Trends in Cognitive Science*, 11:502–503, 2007.

[28] Jose del R. Millan, Frederic Renkens, Josep Mourino, and Wulfram Gerstner. Noninvasive brain-actuated control of a mobile robot by human eeg. *IEEE Transactions on Biomedical Engineering*, 51(6):1026–1033, June 2004.

[29] Arnaud Delorme and S. Makeig. Eeglab: an open source toolbox for the analysis of single-trial eeg dynamics including independent components analysis. *Journal of Neuroscience Methods*, 134:9–21, 2004.

[30] R. Desimone and L. Ungerleider. Mutiple visual areas in the caudal superior temporal sulcus of the macaque. *Journal of Comparative Neurology*, 248:164–189, 1986.

[31] T. Ditman, P. Holcomb, and G. Kuperberg. An investigation of concurrent erp and self-paced reading methodologies. *Psychophysiology*, 44:927–935, 2007.

[32] P.L. Ditunno and V.A. Mann. Right hemisphere specialization for mental rotation in normals and brain damaged subjects. *Cortex*, 26(2):177–188, June 1990.

[33] G. Dogil, H. Ackerman, W. Grodd, H. Haider, H. Kamp, J. Mayer, A. Reicker, and D. Wildgruber. The speaking brain: a tutorial introduction to fmri experiments in the production of speech, prosody, and syntax. *Journal of Neurolinguistics*, 15(1): 59–90, 2002.

[34] E. Donchin, K. M. Spencer, and R. Wijesinghe. The mental prosthesis: assessing the speed of a p300-based brain-computer interface. *IEEE Transactions on Rehabilitation Engineering*, 8:174–179, 2000.

[35] M. D'Zmura, S. Deng, T. Lappas, S. Thorpe, and Ramesh Srinivasan. Toward eeg sensing of imagined speech. In *Proceedings of the 2009 International Human-Computer Interaction Conference (HCII)*, 2009.

[36] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *The Annals of Statistics*, 32:407–451, 2004.

[37] F. Esposito, C. Mulert, and R. Goebel. Combined distributed source and single-trial eeg-fmri modeling: Application to effortful decision making processes. *Neuroimage*, 47:112–121, 2009.

[38] Derry FitzGerald, Eugene Coyle, and Bob Lawlor. Independent subspace analysis using locally linear embedding. In *Proceedings of the Sixth International Conference on Digital Audio Effects*, 2003.

[39] Packard Center for ALS Research. Als statistics. July, 2009. `http://www.alscenter.org`.

[40] Angela Friederici. Syntactic, prosodic, and semantic processes in the brain: Evidence from event-related neuroimaging. *Psycholinguistic research*, 30:237–250, 2001.

[41] Angela Friederici. Towards a neural basis of auditory sentence processing. *Trends in Cognitive Science*, 6:78–84, 2002.

[42] Angela Friederici, B. Opitz, and D. Yves von Cramon. Segregating semantic and syntactic aspects of processing in the human brain: an fmri investigation of different word types. *Cerebral Cortex*, 10:698–705, 2000.

[43] K. Friston. Experimental design and statistical parametric mapping. In Frackowiak et al., editor, *Human brain function*. Academic Press, 2nd edition, 2003.

[44] I. Gauthier, W. Hayward, M. Tarr, A. Anderson, P. Skudlarski, and J. Gore. Bold activity during mental rotation and viewpoint-dependent object recognition. *Neuron*, 34:161–171, March 2002. brainmap search 30291.

[45] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, and H. Stanley. Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101: e215–e220, 2000.

[46] J. Gore, S. Horovitz, C. Cannistraci, and Skudlarskim P. Integration of fmri, nirot, and erp for studies of human brain function. *Magnetic Resonance Imaging*, 24:507–513, 2006.

[47] G. Gratton, M.G. Coles, and E. Donchin. A new method for off-line removal of ocular artifact. *Electoencephalography and Clinical Neurophysiology*, 55:468–484, 1983.

[48] C. Grefkes and G. Fink. The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, 207:3–17, 2005.

[49] A. Haig, E. Gordon, G. Rogers, and J. Anderson. Classification of single-trial erp subtypes: application of globally optimal vector quantization using simulated annealing. *Electoencephalography and Clinical Neurophysiology*, 94:288–297, 1995.

[50] I. Harris, G. Egan, C. Sonkkila, H. Tochon-Danguy, G. Paxinos, and J. Watson. Selective right parietal lobe activation during mental rotation: a parametric pet study. *Brain*, 123:65–73, 2000.

[51] Ping He, Glenn Wilson, Christopher Russell, and Maria Gerschutz. Removal of ocular artifacts from the eeg: a comparison between time-domain regression method and adaptive filtering method using simulated data. *Medical and Biological Engineering and Computing*, 45:495–503, 2007.

[52] Martin Heil. The functional significance of erp effects during mental rotation. *Psychophysiology*, 39(5):535–545, September 2002.

[53] Martin Heil and Petra Jansen-Osmann. Children's left parietal brain activation during mental rotation is reliable as well as specific. *Cognitive Development*, 22(2):280–288, June 2007.

[54] M. Hurd. Functional neuroimaging motor study.

[55] S. J. Hutt and Hugh Fairweather. Information processing during two types of eeg activity. *Electroencephalography and clinical neurophysiology*, 39(1):43–51, July 1975.

[56] A. Hyvarinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10:626–634, 1999.

[57] E. C. Ifeachor, B. W. Jervis, E. L. Morris, E. M. Allen, and N. R. Hudson. New online method for removing ocular artefacts from eeg signals. *Medical and Biological Engineering and Computing*, 24(4):356–364, July 1986.

[58] P. Indefrey and W. Levelt. The spatial and temporal signatures of word production components. *Cognition*, 92:101–144, 2004.

[59] K. Jordan, T. Wustenberg, H. Heinze, M Peters, and L Jancke. Women and men exhibit differnt cortical activation patterns during mental rotation tasks. *Neuropsychologia*, 40:2397–2408, December 2002.

[60] Oliver Josephs, Robert Turner, and Karl Friston. Event-related fmri. *Human Brain Mapping*, 5(4):243–248, 1997.

[61] T. P. Jung, S. Makeig, and J. Westerfield. Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping*, 14:166, 2001.

[62] Tzyy-Ping Jung, Scott Makeig, Colin Humphries, Te-Won Lee, Martin McKeown, Vicente Iragui, and Terrence Sejnowski. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37:163–178, 2000.

[63] Alexander Kaplan. The problem of segmental description of human electroencephalogram. *Human Physiology*, 25(1):107–114, 1999.

[64] Simon P. Kelly, Edmund C. Lalor, Richard B. Reilly, and John J. Foxe. Visual spatial attention tracking using high-density ssvep data for independent brain-computer interaction. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(2):172–178, June 2005.

[65] J. Kenemans, P. Molenaar, M. Verbaten, and Slangen J. Removal of ocular artifact from the eeg: a comparison of time and frequency domain methods with simulated and real data. *Psychophysiology*, 28(1):114–121, 1991.

[66] P. Kennedy, R. Bakay, M. Moore, K. Adams, and J. Goldthwaite. Direct control of a computer from the human central nervous system. *IEEE Transactions on Rehabilitation Engineering*, 8(2):198–202, June 2000.

[67] James N. Knight. Signal fraction analysis and artifact removal in eeg. Master's thesis, Colorado State University, 2003.

[68] J. Kohlmorgen and B. Blankertz. A simple generative model for single-trial eeg classification. In *Proceedings of the International Conference on Artificial Neural Networks*, 2002.

[69] S. Kosslyn. *Image and brain: the resolution of the imagery debate*. MIT Press, 1994.

[70] M. Kurtas and S. Hillyard. Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207:203–205, 1980.

[71] J. Langlois, W. Rutland-Brown, and M. Wald. The epidemiology and impact of traumatic brain injury: a brief overview. *Journal of Head Trauma Rehabilitation*, 21: 375–378, 2006.

[72] E. Learned-Miller and J. Fisher. Ica using spacings estimates of entropy. *Journal of Machine Learning Research*, 4:1271–1295, 2003.

[73] P. Locher and C. Nodine. Influence of stimulus symmetry on visual scanning patterns. *Perception and Psychophysics*, 13(3):408–412, 1973.

[74] P. Manoilov. Eeg power spectrum analysis during mental task performance. In *Proceedings of the International Conference on Computer Systems and Technologies*, 2006.

[75] R. Mappus, R. Ferguson, K. Czechowski, and P. Corballis. Spotting differences: how qualitative asymmetries influence visual search. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, 2005.

[76] R. Mappus, P. Corballis, and M. Jackson. Enhancing brain-machine interface throughput using simultaneous activation. In *ACM SIGCHI: Works-in-progress*, 2009.

[77] D. McFarland, W. Sarnacki, and J. Wolpaw. Brain-computer interface (bci) operation: optimizing information transfer rates. *Biological Psychology*, 63:237–251, 2003.

[78] C. McGibbon, T. Palmer, D. Goldvasser, and D. Krebs. Kalman filter detection of blinks in video-oculography: applications for vvor measurement during locomotion. *Journal of Neuroscience Methods*, 106:171–178, 2001.

[79] M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski. Analysis of fmri data by blind separation into independent spatial components. *Human Brain Mapping*, 6:160–188, 1998.

[80] H. Mohseni, E. Wilding, and S. Sanei. Preprocessing of event-related potential signals via kalman filtering and smoothing. In *Proceedings of the 15th international conference on digital signal processing*, 2007.

[81] Suresh D. Muthukumaraswamy, Blake W. Johnson, and Jeff P. Hamm. A high density erp comparison of mental rotation and mental size transformation. *Brain and Cognition*, 52:271–280, 2003.

[82] N. Neumann, T. Hinterberger, J. Kaiser, U. Leins, Niels Birbaumer, and A. Kubler. Automatic processing of self-regulation of slow cortical potentials: evidence from brain-computer communication in paralysed patients. *Clinical Neurophysiology*, 115 (3):628–635, 2004.

[83] N. Nicolaou, S. Nasuto, and J. Geogiou. Single-trial event-related potential analysis for brain-computer interfaces. In *Proceedings of the 2008 Convention of Society for the study of artificial intelligence and the simulation of behavior*, 2008.

[84] B. Obermaier, C. Guger, C. Neuper, and G. Pfurtscheller. Hidden markov models for online classification of single trial eeg data. *Pattern Recognition Letters*, 22:1299–1309, 2001.

[85] F. Oveisi. Eeg signal classification using nonlinear independent component analysis. In *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2009.

[86] Stephen Palmer and Kathleen Hemenway. Orientation and symmetry: Effects of multiple, rotational, and near symmetries. *Journal of Experimental Psychology*, 4(4): 691–702, 1978.

[87] Stephen Palmer, Joseph Brooks, and Rolf Nelson. When does grouping happen? *Acta Psychologica*, 114(3):311–330, November 2003.

[88] W. Penny and S. Roberts. Dynamic models for nonstationary signal segmentation. *Computers and biomedical research*, 32:483–502, 1999.

[89] F. Peronnet and M.J. Farah. Mental rotation: an event-related potential study with validated mental rotation task. *Brain and Cognition*, 9(2):279–88, March 1989.

[90] B. O. Peters, G. Pfurtscheller, and H. Flyvbjerg. Automatic differentiation of multichannel eeg signals. *IEEE Transactions on Biomedical Engineering*, 48(1):111–116, January 2001.

[91] Gert Pfurtscheller, D. Flotzinger, W. Pregenzer, Jonathan R. Wolpaw, and D. J. McFarland. Eeg-based brain computer interface (bci): search for optimal electrode positions and frequency components. *Medical Programming Technology*, 21:111–121, 1996.

[92] K. Podzebenko, G. Egan, and J. Watson. Real and imaginary rotary motion processing: functional parcellation of the human parietal lobe revealed by fmri. *Journal of Cognitive Neuroscience*, 17(1):24–36, January 2005. doi: 10.1162/0898929052879996.

[93] B. R. Postle, J. S. Berger, A. M. Taich, and Mark D'Esposito. Activity in human frontal cortex associated with spatial working memory and saccadic behavior. *Journal of Cognitive Neuroscience*, 12 Supp. 2:2–14, 2000.

[94] M. Potter, N. Gadhok, and W. Kinser. Separation performance of ica on simulated eeg and ecg signals contaminated by noise. In IEEE, editor, *IEEE Canadian Conference on Electrical and Computer Engineering*, volume 2, pages 1099–1104. IEEE, 2002.

[95] Martin Pregenzer and Gert Pfurtscheller. Frequency component selection for an eeg-based brain to computer interface. *IEEE Transactions on Rehabilitation Engineering*, 7(4):413–419, December 1999.

[96] Adriane Randolph. *Individual-technology fit: matching individual characteristics and features of biometric interface technologies with performance.* PhD thesis, Georgia State University, 2007.

[97] Sam Roweis and Lawrence Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, December 2000.

[98] H. Sakata, M. Taira, M. Kusunoki, A. Murata, K. Tsutsui, Y. Tanaka, W. Shein, and Y. Miyashita. Neural representation of three-dimensional features of manipulation objects with stereopsis. *Journal of Experimental Brain Research*, 128:160–169, 1999.

[99] G. Schalk, D. McFarland, T. Hinterberger, N. Birbaumer, and J. Wolpaw. Bci2000: A general-purpose brain-computer interface (bci) system. *IEEE Transactions on Biomedical Engineering*, 51:1034–1043, 2004.

[100] Roger Shepard and Jacqueline Metzler. Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703, February 1971.

[101] Sameer Singh. Eeg data classification with localised structural information. In *proceedings of the 15th intenational conference on pattern recognition*, pages 271–274. IEEE, 2000.

[102] C. Summers, B. Ivins, and K. Schwab. Traumatic brain injury in the united states: an epidemiologic overview. *Mount Sinai Journal of Medicine*, 76:105–110, 2009.

[103] A. Tang, M. Sutherland, C. McKinney, J. Liu, Y. Wang, L. Parra, A. Gerson, and P. Sajda. Classifiying single-trial erps from visual and frontal cortex during free viewing. In *IEEE Proceedings of the 2006 International Joint Conference on Neural Networks*, 2006.

[104] J.B. Tenenbaum, V. de Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.

[105] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, 58(1):267–288, 1996.

[106] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3:71–87, 1991.

[107] P. A. van der Helm and E. Leeuwenberg. Accessibility: a criterion for regularity and hierarchy in visual pattern codes. *Journal of Mathematical Psychology*, 35:151–213, 1991. URL `http://www.nici.kun.nl/ peterh/publications.html`.

[108] P. A. van der Helm, R. van Lier, and E. Leeuwenberg. Serial pattern complexity: irregularity and hierarchy. *Perception*, 21:517–544, 1992. URL `http://www.nici.kun.nl/ peterh/publications.html`.

[109] J Vanrie, E. Beatse, J. Wagemans, S. Sunaert, and P. Van Hecke. Mental rotation versus invariant features in object perception from different viewpoints: an fmri study. *Neuropsychologia*, 40:917–930, 2002.

[110] R. Vilimek and T. Zander. *Universal access in human-computer interaction: intelligent and ubiquitous interaction environments*, chapter BC(eye): Combining eye-gaze input with brain-computer interaction, pages 593–602. Springer, 2009.

[111] G. Vingerhoets, F. de Lange, P. Vandemaele, K. Deblaere, and E. Achten. Motor imagery in mental rotation: an fmri study. *NeuroImage*, 17:1623–1633, November 2002.

[112] M. Von Spreckelsen and B. Bromm. Estimation of single-evoked cerebral potentials by means of parametric modeling and kalman filtering. *IEEE Transactions on Biomedical Engineering*, 35:691–700, 1988.

[113] Sergiy Vorobyov and Andrzej Cichocki. Blind noise reduction for multisensory signals using ica and subspace filtering, with application to eeg analysis. *Biological Cybernetics*, 86:293–303, 2002.

[114] J. Wagemans. Detection of visual symmetries. *Spatial Vision*, 9(1):9–32, 1995.

[115] Y. Wang, M. Sutherland, L. Sanfratello, and A. Tang. Single-trial classification of erps using second-order blind identification (sobi). In *IEEE Proceedings of the third international conference on machine learning and cybernetics*, 2004.

[116] J. Weston, A. Elisseff, B. Scholkopf, and M. Tipping. Use of the zero-norm with linear models and kernel methods. *Journal of machine learning research*, 3:1439–1461, 2003.

[117] Jonathan R. Wolpaw, Niels Birbaumer, Dennis J. McFarland, Gert Pfurtscheller, and Theresa M. Vaughan. Brain-computer interfaces for communication and control. *Clinical Neurophysiology*, 113:767–791, 2002.

[118] M. Wraga, S. Creem, and D. Proffitt. Updating displays after imagined object and viewer rotations. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 26:151–168, 2000.

[119] H Zabrodsky, S. Peleg, and D. Avnir. A measure of symmetry based on shape similarity. In *Proceedings of IEEE Computer Society on Computer Vision and Pattern Recognition*, 1992.

[120] Xiaoyuan Zhu, Cuntai Guan, Jiankang Wu, Yimin Cheng, and Yixiao Wang. Bayesian method for continuous control in eeg-based brain-computer interface. In *Proceedings of the 27th Annual International Conference of Engineering in Medicine and Biology Society*, 2005.

[121] Xiaoyuan Zhu, Cuntai Guan, Jiankang Wu, Yimin Cheng, and Yixiao Wang. Expectation-maximization method for eeg-based continuous cursor control. *EURASIP Journal on Applied Signal Processing*, 2007:26–36, 2007.

[122] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, 67:301–320, 2005.

[123] H. Zou, T. Hastie, and R. Tibshirani. Sparse principal components analysis. *Journal of computational and graphical statistics*, 15:265–286, 2006.